VE-STATE Development Approach

Brian Gregor

7/9/18

This memo describes the planned approach to developing VisionEval modules that can work with modules already developed for VE-RSPM to implement the GreenSTEP model in the VisionEval model system. This memo focuses on the following 2 development aspects:

1. How VE-RSPM modules will be tested and modified to confirm that they will work in a statewide application; and,
2. How new modules will be developed to synthesize Bzone characteristics from Azone and Marea characteristics needed to provide datasets used by other modules.

The memo does not address full testing of VE-STATE modules with a state dataset. That will be done in a later memo.

Testing VE-RSPM modules that will be used in VE-STATE

It is anticipated that most of the modules developed for VE-RSPM can be used in VE-STATE but they need to be tested to confirm that they do and revised if they don't. The testing that was done during the development of VE-RSPM used datasets developed for the Rogue Valley Metropolitan Planning Organization (RVMPO) planning area. These datasets have only one Marea and one Azone. Since VE-STATE models will have several Mareas and several Azones the first test to be done to determine whether VE-RSPM modules will work in VE-STATE is to run the modules with datasets having several Mareas and Azones. This can be done by duplicating the records in the datasets several times and applying new zone names so that there are several Azones and Mareas that have identical values. An R script will be written to automate the process. VE-RSPM will then be run using the modified inputs. Corrections will be made to any modules that are found not to work correctly with datasets for several Azones and Mareas.

Developing modules to synthesize Bzones and their attributes

The main difference between VE-RSPM models and VE-STATE models is that a number of VE-RSPM inputs are specified at the Bzone level. Examples include numbers of dwelling units by type and numbers of jobs by sector. VE-STATE models run at a higher level of abstraction than VE-RSPM models and don't have Bzone level inputs.

Bzone level attributes are required by a number of modules so methods need to be developed for synthesizing a representative set of Bzones and their characteristics from policies and attributes specified at the Azone and Marea levels. Something like this is done in the GreenSTEP model where a likely distribution of neighborhood population density is synthesized from the overall metropolitan area density. Azone level inputs are provided for base year population and area by development type (metropolitan, town, rural), population growth by development type, and the ratio of urban area growth to population growth. From these inputs, average density is calculated by Azone and development type and a model is applied to synthesize a distribution of neighborhood densities from the average density.

Attachment A includes an excerpt from the GreenSTEP documentation that explains the method and model in detail.

All of the VE-RSPM modules which assign Bzone characteristics are contained in the VELandUse package. The modules that are developed to synthesize Bzones and their characteristics will be placed in a VESimLandUse package. When a VE-STATE model is run, the modules in the VESimLandUse package will be run instead of the modules in the VELandUse package. Otherwise the model setup will be almost the same for VE-STATE and VE-RSPM.[1] The use of the VESimLandUse modules will not be limited to statewide applications however. Users could simulate Bzones in an RSPM-type application that would enable metropolitan area planners to more easily define and model alternative land use scenarios as is done in RPAT (Rapid Policy Analysis Tool) applications.

Required Bzone Attributes

Land use modeling in VE-RSPM is more complex than in the RSPM and GreenSTEP. There are two reasons for this. First, modules were designed to produce datasets needed to run the new multi-modal travel model. This module require several activity density, diversity (i.e. activity mixing), and destination accessibility measures that in turn require households and employment to be located at the Bzone level to calculate those measures. In addition, a few multi-modal network and service level measures need to be calculated. Second, locating jobs at the Bzone level opened up opportunities for significant improvements to the travel demand management (TDM) and parking pricing modules to establish more realistic relationships between policies and the households they would affect. It was a simple extension to locate household workers at job sites and then use that information of translate job site TDM and parking policies back to households. Finally, an aspect of VE-RSPM that is more complicated than GreenSTEP, but not more complicated than RSPM, is how simulated households are assigned to Bzones. In VE-RSPM a numbers of single family and multifamily dwelling units are assigned to Bzones as inputs along with the relative income distribution of households in each Bzone. VE-RSPM models the housing choice of each household based on the overall supply of housing of each type in the Azone and the household characteristics. Then the model assigns each household to a Bzone based on the household's housing choice and income, and on the relative supplies of housing of the type and household income distribution in the Bzone.

The Bzone attributes that need to be synthesized are:

- Destination accessibility (i.e. accessibility to jobs and housing) measured consistent as it is used in the multi-modal travel model – This information is one of the 5D measured used in the new RSPM multi-modal travel model. In the RSPM it is calculated from numbers of households and jobs and zone centroid locations. Since synthetic zones won't have physical locations, it can't be calculated simply from households and employment by zone.
- Number of households proportional split of dwelling units between single family and multifamily – Number of households and dwelling unit split by Bzone is needed in order to assign households to Bzones.

---

[1] One other difference is that VE-STATE will also more likely run an optional module to calculate added road use taxes to pay for assumed road improvements.

- Number of jobs by sector (retail, service, other) is used in calculating several diversity measures used in the RSPM multi-modal travel model. The number of jobs is also used to associate household workers with workplace Bzones.
- Area type and development type – Some practical system of zonal development classification is needed for organizing policy inputs. In GreenSTEP, 3 development types are used. These are called metropolitan, town, and rural. It is proposed that a classification system merge these development types with the place type system used in RPAT where place types are defined as a combination of area types and development types (more on this below). Policies such as travel demand management policies will be specified by Azone and area type and/or development type. These designations will also be used in the calculation of the design and distance to transit '5D' measure categories that are used in the RSPM multi-modal travel model.

Proposed Approach for Synthesizing Bzones and their Attributes

Following is an outline of the proposed zone synthesis process in approximate order.

1. The user provides inputs on:
    - Azone proportional split of dwelling units by location type[2] (metropolitan, town and rural)
    - Azone proportional split of workers by job site location types (For example proportions of rural residents in the Azone who work in rural locations, town locations, or the metropolitan area)
    - Marea proportional split of Marea employment among Azones.
2. Total activity – numbers of households and jobs – will determine the number of SimBzones in the Azone. SimBzones will have equal amounts of activity and unequal areas since activity density will vary among SimBzones. An appropriate average SimBzone activity value will be determined through evaluation of the Smart Location Database (SLD).
3. Models will be applied to select a destination accessibility value for each Bzone. For metropolitan type development, the model will create a distribution of destination accessibility values that is consistent with the overall activity density in the metropolitan area. Random sampling from the distribution will be used to assign destination accessibility values to metropolitan SimBzones. This model will be similar in nature to the density model in GreenSTEP. Models will also be developed for town and rural types, but more investigation is needed in order to determine their form.
4. Activity density of each SimBzones will be determined as a function of the destination accessibility of each zone (because destination accessibility is a measure of activity density at a larger geographic scale). A model will be estimated from the SLD which creates a distribution of zone densities as a function of destination accessibility. An adjustment process, such as an iterative proportional fitting process (IPF), will be used to adjust densities and destination accessibilities so that the overall activity density of all the zones in a metropolitan area is equal to the input value.
5. Further subdivision of the metropolitan area into area types will be done as a function of the destination accessibility and activity density of each zone. 4 such area types are proposed which

---

[2] These location types are equivalent to the development types of the GreenSTEP and EERPAT models.

meld the RPAT types with GreenSTEP/EERPAT location types: urban core, close in community, and suburban/town, low density/rural. The final typology and the relationship of area types to destination accessibility and activity density will be developed through examination of the SLD dataset. It is envisioned that area types will be defined as fuzzy sets rather than crisp sets. Although some SimBzones may be wholly one type, many SimBzones will have degrees of membership in several types. Using fuzzy sets is a more realistic recognition of the nature of area types and avoids aberrations resulting from threshold effects.

6. The total activity in each SimBzone is split into households and jobs using a model which relates zonal mixing to destination accessibility, activity density, and area type. This model will be specified and estimated based on investigations using the SLD data. It is anticipated that the model will produce distributions of activity splits from which values will be drawn. IPF or some other adjustment process will be used to adjust values so that the aggregation of the splits for all the SimBzones in an Azone is consistent with the Azone inputs. This model will also need to specify the split of land area between households and jobs.
7. Once the number of households is determined for each SimBzone, the split of dwelling units by housing type (single family, multifamily) as a function of activity density. The SLD and census data will be used to develop this model. Earlier work on developing such a model for GreenSTEP has been done which will be consulted. IPF or some other adjustment process will be used to fit the distribution of SimBzone values to Azone level control totals. This will allow users to specify Azone ratios as policy inputs.
8. A variant of the VE-RSPM housing model will be applied to assign households to housing types and to Bzones. One thing to be worked out is whether the allocation to SimBzones considers household income or not. In the RSPM, relative Bzone income distributions are input and these are used in the process of allocating households to Bzones. Something similar could be done by Azone and area type for VE-STATE. This would enable VE-STATE users to model general relationships of income to parts of the metropolitan area (e.g. the effect of gentrification in the urban core).
9. Jobs in each SimBzone will be split into numbers of retail, service, and other jobs. The approach for doing this is yet to be determined. SimBzone splits will be constrained so that they total to Azone splits that are inputs. The model will probably be a function of the destination accessibility, employment density, and mixing of households and employment in the SimBzone. The SLD data will be used to develop and estimate this model. Thought will be given as to whether there should be control totals on the mix for a metropolitan area. If so, IPF or some other adjustment process would need to be used to match the totals.
10. Workers will be assigned to SimBzone job sites. How this is done is yet to be determined. It is proposed that an agile development approach be used where the first iteration of the model will be a random assignment of workers to job sites. Other extensions that could be considered if there is time/budget investigation could be done using LEHD data and SLD data to look for relationships between worker residence by area type and worker job site by area type. Relative income could also be considered.
11. Once numbers of households and numbers of jobs by type are assigned to SimBzones, all the remaining density and diversity measures can be calculated.
12. The distance to transit measure will be modeled for metropolitan SimBzones as a function of the metropolitan-level transit supply measure and the SimBzone attributes for destination

accessibility, density, and mixing. A combination of National Transit Database (NTD) and SLD data will be used to develop this model. The structure is to be determined.

13. Development type, like the RPAT development types (e.g. residential, employment, mixed, transit-oriented development, greenfield), will be assigned to SimBzones based on the density, diversity, and distance to transit measures. These development types, like the area types may be fuzzy sets. The SLD will be used to create the development type specifications. The design will enable model users to input Marea goals for the proportional split of development types. The model will adjust the allocation of development types consistent with the goals but constrained to plausible levels.
14. Network design measures used by the RSPM multi-modal travel model (e.g. multi-modal network density, pedestrian network density) will be applied based on inputs related to area and development type. The SLD will be used to identify ranges of values by area and development type. Users will specify in inputs goals relative to these ranges by Azone, development type and area type.
15. The parking pricing, travel demand management, and car service inputs will be specified by Azone, area type, and development type. These will then be translated to the SimBzone based on the SimBzone area type and development type. After that is done, the AssignDemandManagement, AssignParkingRestrictions, and AssignCarSvcAvailability modules can run as they currently do. Thought will be given as to how to simplify inputs so that users are not required to provide inputs for every combination of Azone, area type and development type.

Appendix A: GreenSTEP method for synthesizing neighborhood density

Several land use characteristics must be predicted for households in order to estimate household vehicle ownership and vehicle travel. These include the type of area where the household resides (metropolitan, town, rural), the population density (persons per square mile) of the Census tract where the household resides, and the urban form characteristics of the Census tract where the household resides (urban mixed-use vs. other). Although the vehicle and travel models require Census tract level characteristics, this level of geography is not explicitly represented in the model because GreenSTEP was developed to model GHG mitigation policies at a statewide or metropolitan level. Therefore, models and calculation methods were developed to estimate likely Census tract characteristics for urban areas based on larger scale characteristics. The GreenSTEP and RSPM models take different approaches to model these characteristics.

In GreenSTEP, land use characteristics are assigned to households in the following steps:

1. Each household in each county is assigned to one of three development types - metropolitan, town, or rural. The metropolitan development type includes urbanized portions of a metropolitan statistical area. In Oregon they are the portions of a metropolitan area that are within the metropolitan urban growth boundary. The town development type areas within the urban growth boundaries of incorporated cities that are not within a metropolitan area urban growth boundary and areas within the boundary of an urban unincorporated community that is not within a metropolitan area urban growth boundary. The rural development type includes all other places outside of urban growth boundaries.
2. The geographic extent of urban growth in metropolitan and other urban areas in each county or metropolitan district is calculated.
3. The overall average density for each development type is calculated.
4. Households are assigned a neighborhood population density (i.e. Census tract population density) as a function of the overall metropolitan, town or rural area density where it is located.
5. Households in metropolitan areas are designated as being in an urban mixed-use community/neighborhood or not, based on Census tract density and metropolitan goals for urban mixed-use development.

Households are assigned to metropolitan, town, and rural development types based on 1) the base year distribution of population by development type by county or metropolitan district and, 2) forecasts or assumptions about the proportions of future population growth by type. The base year distribution is developed from Census data, using Census tract population density as an indicator or from metropolitan area household inventories. From these data, the total proportions of households to be assigned to each development type are computed. Households are then randomly assigned to each type using the calculated proportions as probabilities. Since the forecasts of population growth proportions are inputs to the model, they can be modified to test the effects of alternative land use policies on vehicle travel (e.g. what is the effect of a greater proportion of population growth occurring in rural areas).

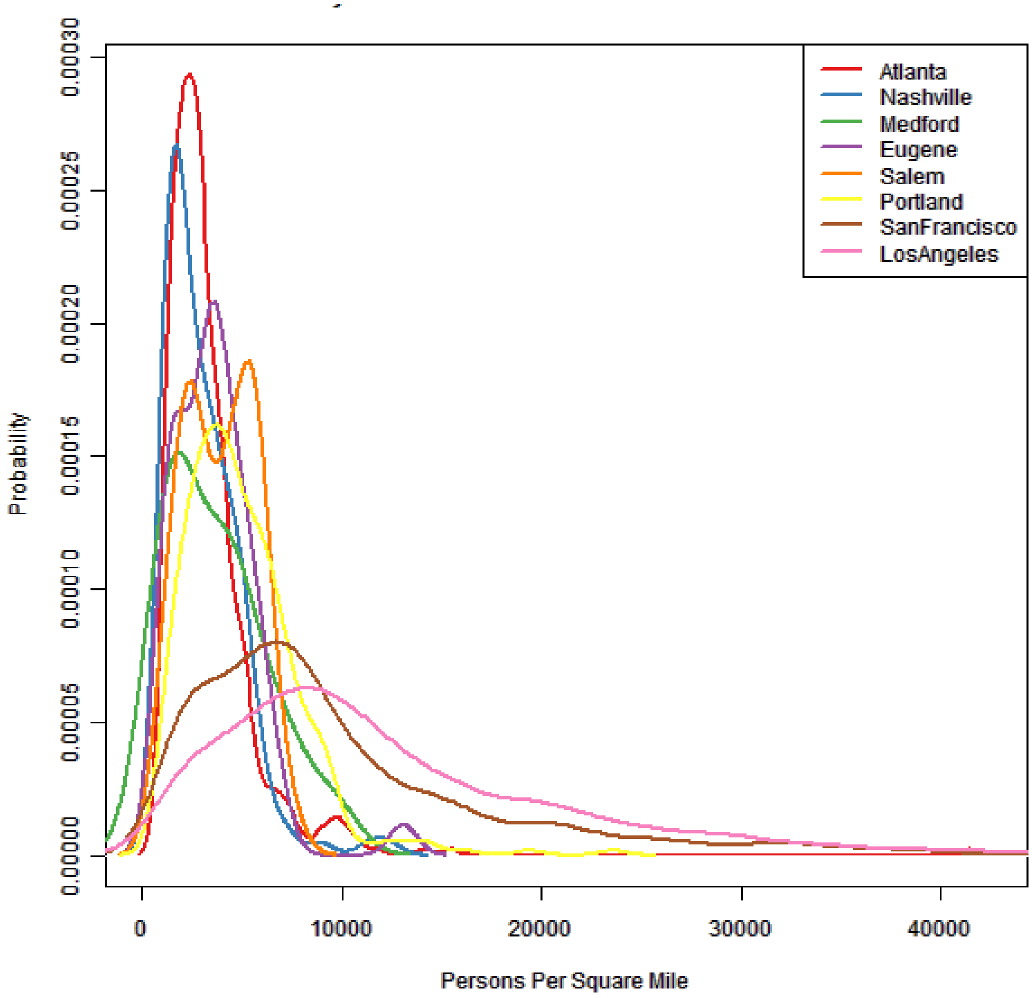
The geographic extent of metropolitan and other urban areas is calculated from base year measurements of urban growth boundary areas by type for each county or metropolitan district and policy inputs which describe how rapidly urban growth boundaries may grow relative to population. For example, a value of 0.5 for metropolitan development means that the urban growth area for the metropolitan area will grow at half the rate of metropolitan population growth.

Households are assigned a neighborhood population density based on the development type that they occupy. A uniform population density is assumed for the rural portions of each county or district. Although densities in rural areas vary, the degree of variation is not large and the variation tends to be localized. For the base year, this density is the household weighted average of rural Census tract densities. This weighting approach is more appropriate than an area-weighted average, which would underestimate average densities because of the large undeveloped areas of public land holdings in most Oregon counties. For forecast years, it is assumed that additional rural population will be added at a density of one household per two acres, since that is the minimum size allowed by state rules for new rural development. The new development is averaged with the base year density to arrive at the forecast rural density.

For metropolitan and town development types, neighborhood population densities are calculated from the overall population density for the area that is urbanized or within an urban growth boundary. The overall population density for one of these types is computed from the number of people assigned to the development type, and the total urban area computed for the type. For households assigned to the town development type, their neighborhood density is the same as the overall average density for the town since small cities tend to be composed of few Census tracts and population densities in small cities tend to be fairly uniform. This is not done for the metropolitan type households.

The assumption of uniform density is not valid for metropolitan areas since Census tract densities can vary by orders of magnitude within a metropolitan area. This can be seen in Figure 9, which compares the Census tract population density distributions of a number of U.S. metropolitan areas.

FIGURE 1. POPULATION DENSITY DISTRIBUTIONS FOR SELECTED METROPOLITAN AREAS



Previous versions of GreenSTEP's census tract density model used the observed census tract density distributions for the Atlanta, Portland, San Francisco and Los Angeles metropolitan areas as prototype sampling distributions. These were supplemented by several additional prototypes having densities less than Atlanta and greater than Los Angeles. The lower density distributions were synthesized by shifting the Atlanta distribution leftward, while the higher density distributions were synthesized by shifting the Los Angeles distribution rightward. The data for each prototype included the overall average metropolitan density and the distribution of Census tract population densities. The density distribution for any given metropolitan area was determined by interpolating between the density distributions of the prototypes whose average densities bound the average density for the subject metropolitan area.

The previous census tract density model had to be replaced because it would not work in and earlier metropolitan version of GreenSTEP (for the Portland metropolitan area) which needed to assign density for subareas of the metropolitan region. The metropolitan prototype approach would not work because district densities have a larger range than metropolitan densities. A scalable approached had to be developed instead.

The procedure for developing the new model again used the census tract level population and density information for the prototype metropolitan areas (Portland, Los Angeles, San Francisco, and Atlanta). Census tracts having population densities less than 1000 persons per square mile were dropped from the estimation data set as being unrepresentative of urbanized areas. Census tract densities were normalized by computing the natural log of population density for each tract in an urbanized area and dividing by the natural log of the population weighted average density for the urbanized area. Figure 10 shows the probability distributions of normalized census tract population densities for the four metropolitan areas (histograms) and compares those distributions to normal distributions using the mean and standard deviation values of the observed distributions (dashed red lines).

The graphs in Figure 10 show that the distributions of normalized census tract densities are reasonably well approximated by the normal distributions. Table 8 compares the means and standard deviations for these four metropolitan areas and for the 3 mid-sized metropolitan areas in Oregon (Medford, Salem, Eugene). It can be seen that the mean and standard deviation values are all fairly similar to one another. There is less variation among the Oregon metropolitan means and standard deviations. This similarity of normalized distributions suggests a scalable approach to estimating a reasonable distribution of census tract densities for urban areas of different sizes or portions of urban areas. Based on these results, parameter values of 1.02 and 0.07 were chosen for use in the latest statewide and metropolitan area models.

FIGURE 2. COMPARISON OF NORMALIZED POPULATION DENSITY DISTRIBUTIONS FOR THE PORTLAND, ATLANTA, SAN FRANCISCO, AND LOS ANGELES METROPOLITAN AREAS

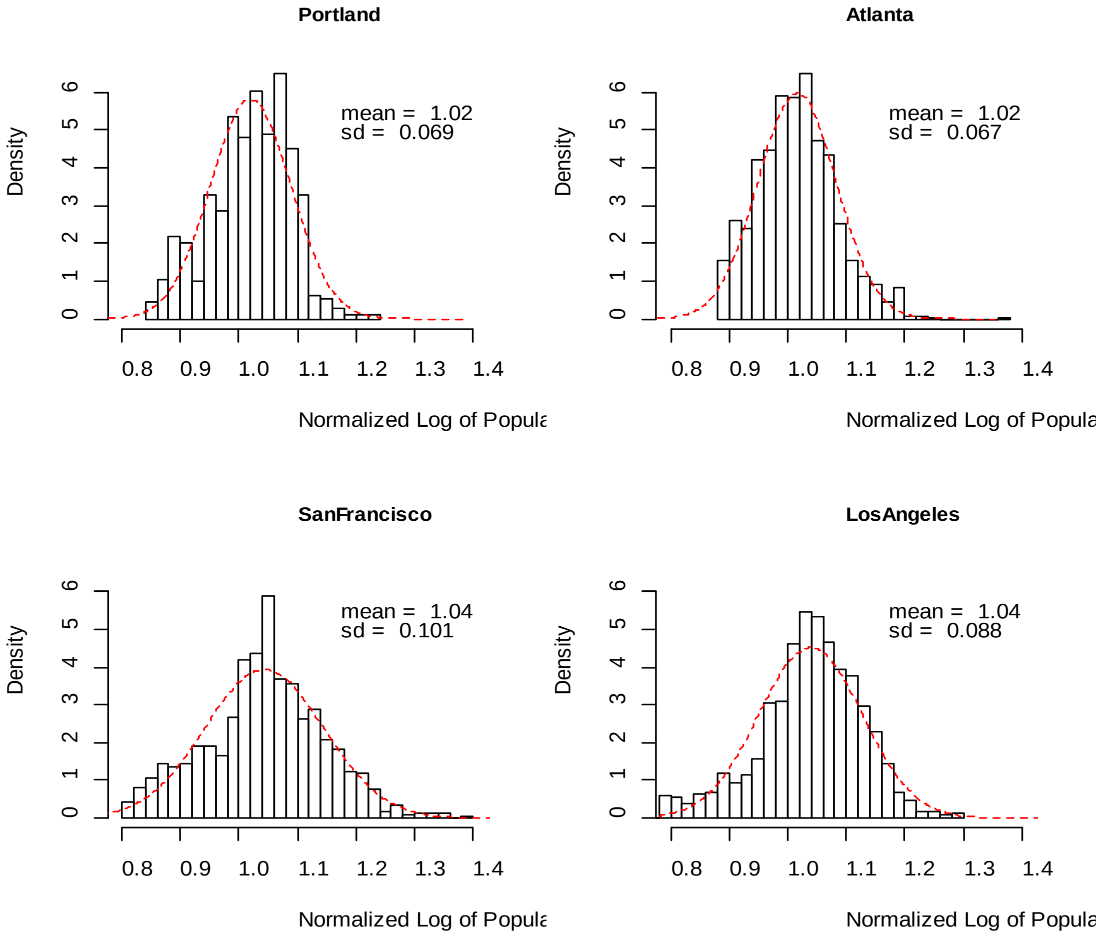


TABLE 1. MEANS AND STANDARD DEVIATION OF NORMALIZED CENSUS TRACT POPULATION DENSITIES FOR SELECTED METROPOLITAN AREAS

Place	Mean	SD
Medford	1.029	0.0882
Salem	1.017	0.0633
Eugene	1.020	0.0714
Portland	1.020	0.0686
Atlanta	1.016	0.0668
San Francisco	1.043	0.1013
Los Angeles	1.039	0.0881

The density model in GreenSTEP to calculate the population density distribution for a metropolitan area or a district within a metropolitan area is as follows:

$$D = \exp(N(\mu, \delta^2) * \log(\mathrm{WtAveDen}))$$

where:

$N(\mu, \delta^2)$ is a normal distribution having mean and standard deviation of the estimated values of 1.02 and 0.07 respectively from which 1 million samples are drawn.

$\mathrm{WtAveDen}$ is the population weighted average density for the metropolitan area or district within the metropolitan area.

Since the population weighted average density is not a known quantity, but the overall average density is, the population weighted average density must be approximated. This can be done because *D* is continuous and average density can be calculated from *D* as follows:

$$\frac{n}{\sum_i^n \frac{1}{d_i}}$$

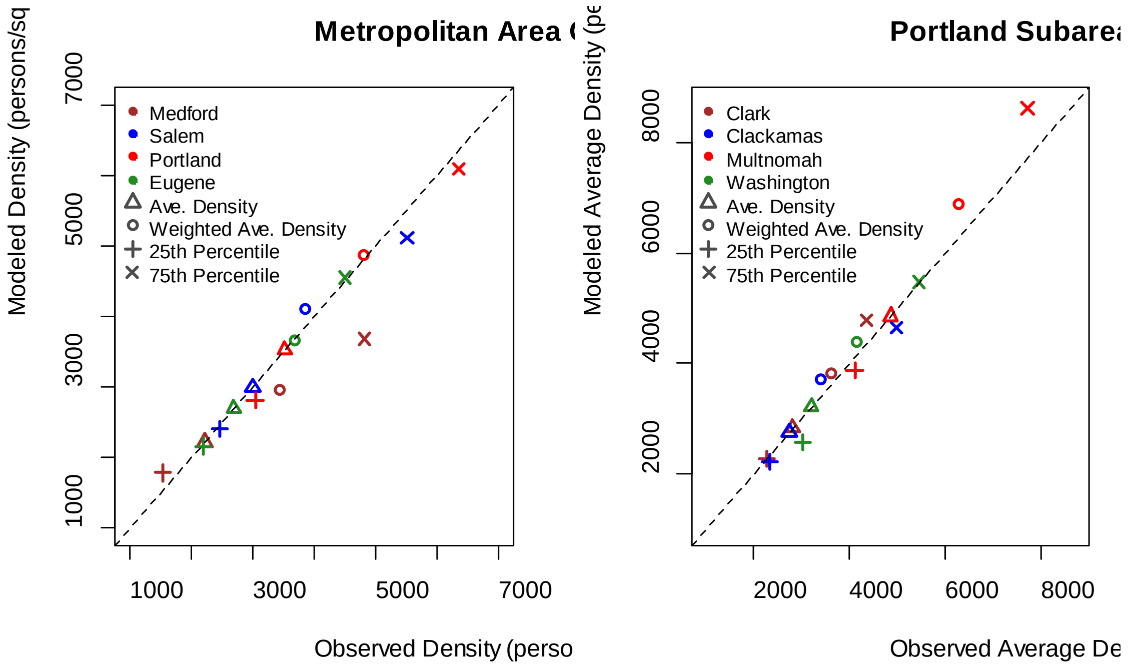
where:
n is the number of samples
d_i is density of each sample

The model uses a binary search algorithm to find a value for the population weighted average density which produces a density distribution whose average density is within 0.1% of the input value.

The scalability of this approach was tested by applying the model to the 4 larger metropolitan areas in Oregon (Portland, Salem, Eugene, Medford) and by applying the model to subareas of the Portland metropolitan area. The Portland subareas were defined by partitioning the census tract dataset by county (Clark, Clackamas, Multnomah, Washington). Figure 11 shows the results for the metropolitan area test in the left-hand graph. The right-hand graph shows the results for the metropolitan subarea test.

The overall average density values are shown by the triangular marks. These all fall on the diagonal line, showing that observed and estimated values are the same, because the model algorithm is designed to match these values. The population weighted average values are shown by the circular marks. These provide an indication of the overall fits of the modeled density distributions to the observed distributions. These are fairly close to the diagonal line for both tests. The vertical cross marks show the 25th percentile density values. These too are close to the diagonal lines. Finally, the angled cross marks show the 75th percentile density values. As might be expected, because the population density distribution has a long right-hand tail, some of these marks depart more from the diagonal line. The model underestimates the 75th percentile census tract density for the Medford metropolitan area and overestimates the 75th percentile census tract density for the Multnomah County subarea. Overall, the modeled distributions do a reasonable job of simulating the observed distributions at different scales.

Figure 3. Comparison of Observed and Modeled Densities for Metropolitan Areas and for Metropolitan Subareas



The current GreenSTEP model assigns households within a County or metropolitan district randomly to development types and to neighborhood densities.