PECAS - for Spatial Economic Modelling

THE ARC PECAS MODEL DEVELOPMENT:
Model Components and Calibration

System Documentation Technical Note

Calgary, Alberta

December 2015

Table of Contents

1. Introduction

The Atlanta Regional Commission's PECAS ("Production Exchange Consumption Allocation System") model began development in the year 2008. It is a spatial economic model for the Atlanta Region in the State of Georgia. Its main purpose is to simulate the future location of activities (industries, households and government), and the development of space by developers, for both forecasting and policy analysis.

The ARC PECAS model includes the two standard PECAS modules: the Activity Allocation module (AA) and the Space Development module (SD). AA follows an aggregate approach and represents how and why industries, government and households choose to locate in different zones or locations in the region. SD follows a microsimulation approach and simulates development at the parcel level, taking into account developers' profit-motivated behavior as well as land and market characteristics. These two modules interact with each other, and both also interact with the Atlanta transport model by providing it with land use data. The transport model, in turn, provides an indication of travel conditions for use in AA.

The base year for the ARC PECAS model is the year 2005 and it has been run through time to 2050. It was originally developed with a zone system involving 78 super districts – also called land use zones (LUZ) – and 2,024 Traffic Analysis Zones (TAZ).

To build the ARC PECAS model, different model components have been developed, including:

- The spatial economic relationships in the region in terms of production and consumption of the categories of goods, services, labor and space (collectively called "Commodities") by industries, households and the government (collectively called "Activities"). This is known as the Aggregate Economic Flow Table (AEFT).
- Rent data for residential and non-residential space.
- The floorspace inventory.
- The parcel database.

To complete the ARC PECAS model development, the two PECAS modules – Activity Allocation and Space Development – were calibrated in several different ways, to set the value of important behavioral parameters.

The objective of this report is to briefly describe the procedures involved in the development of the ARC PECAS model, which include the data collection and processing components, along with the calibration procedures. Most of the procedures involved are contained in spreadsheets, scripts, databases, programing code and other technical documentation. These valuable pieces of information are referred to in this document to indicate the source of the described procedures.

This report is the first of two technical documents reporting the evolution of the ARC PECAS model through time. This first document describes the model development – involving model components and calibration – performed until 2014, in many cases referencing previous documents written during the history of the development of the ARC's PECAS model. This report is not meant to be comprehensive on its own, but rather comprehensive when read in conjunction with the other documents referenced within.

The second report is titled **The ARC PECAS Model Improvements, current use and ABM integration** and covers the recent enhancements and improvements in the ARC PECAS model in order to respond to the current needs of the Atlanta Regional Commission. These include adjustments to work with more recent available data, changes in the model for full integration with the new ABM transport model for Atlanta, and the development and fine-tuning of procedures and data as required to support The Region's Plan Needs Assessment Forecast.

2. Overview of procedure followed to develop the Aggregate Economic Flow Table (AEFT)

The AEFT defines the magnitudes of the interactions (production, consumption, and exchange) among the Activities (model agents) and Commodities (categories of interaction) defined in the ARC PECAS model. This table generally includes the following relationships:

- Industries and accounts producing and consuming (expending on) goods, services, and receipts (for taxes, government funding, investments and capital).
- Industries consuming labor and non-residential space.
- Households providing labor and consuming goods, services, and residential space.

The money flows (including production and consumption) among the 36 activities through the 35 commodities defined for the ARC PECAS model were calculated using data from IMPLAN. The description of the AEFT involving the consumption of space (residential and non-residential) by the activities is addressed in the next section.

The categorizations of the activities and commodities defined for the ARC PECAS model are shown respectively in Tables 1 and 2.

Table 1. Activity categories defined for the ARC PECAS Model

No	Activity Name	Description
Industries		
1	AI01AgMinMan	Agriculture and mining management
2	AI02AgMinProd	Agriculture and mining production
3	AI03ConMan	Construction management
4	AI04ConProd	Construction production
5	AI05MfgMan	Manufacturing management

No	Activity Name	Description
6	AI06MfgProd	Manufacturing production
7	AI07TCUMan	Transport and utilities management
8	AI08TCUProd	Transport and utilities production
9	AI09Whole	Wholesale services
10	AI10Retail	Retail and food services
11	AI11FIRE	Finance insurance and real estate
12	AI12PTSci	Professional and technical services
13	AI13ManServ	Management services
14	AI14PBSOff	Personal and business services office based
15	AI15PBSRet	Personal and business services retail based
16	AI16PSInd	Personal services industrial based
17	AI17Religion	Religious services
18	AI18BSOnsite	Business services onsite
19	AI19PSOnsite	Personal services onsite
20	AI20FedGov	Federal government
21	AI21StLocGov	State and local government services
22	AI22Military	Military
23	AI23GSEdu	Grade school education
24	AI24HiEdu	Higher education
25	AI25Health	Health
Capital Accounts		
26	AA26FedGovAccounts	Federal government accounts
27	AA27StLocGovAccounts	State and local government accounts
28	AA28CapitalAccounts	Capital accounts
Households		
29	AH29HHlt20_12	Households with annual income < 20 thousand and 1 to 2 people
30	AH30HHlt20_3p	Households with annual income < 20 thousand and > 2 people
31	AH31HH2050_12	Households with annual income 20 to 50 thousand and 1 to 2 people
32	AH32HH2050_3p	Households with annual income 20 to 50 thousand and > 2 people
33	AH33HH50100_12	Households with annual income 50 to 100 thousand and 1 to 2 people
34	AH34HH50100_3p	Households with annual income 50 to 100 thousand and > 2 people
35	AH35HHge100_12	Households with annual income > 100 thousand and 1 to 2 people
36	AH36HHge100_3p	Households with annual income > 100 thousand and > 2 people

Table 2. Commodity categories defined for the ARC PECAS Model

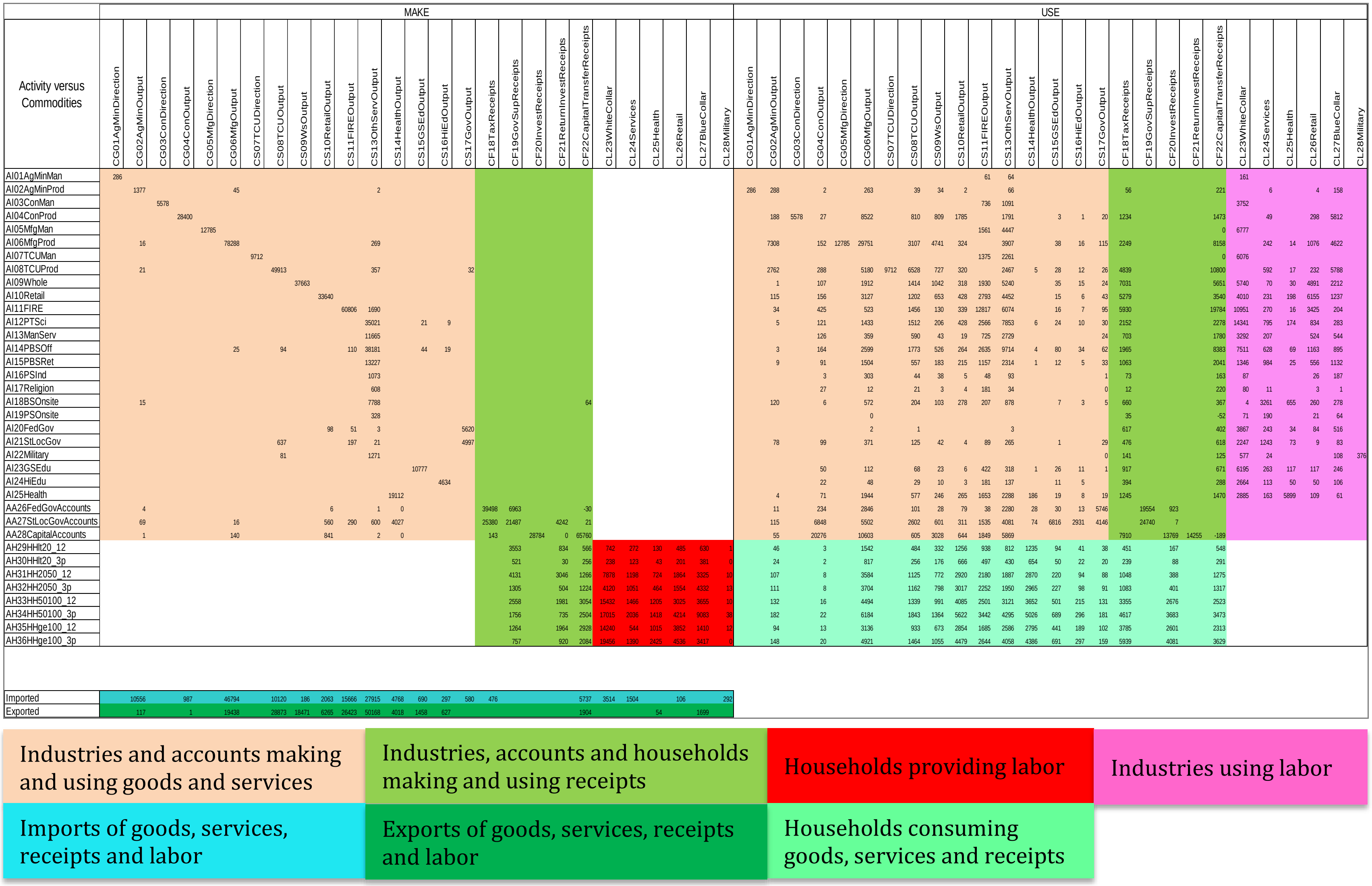
No	Commodity Name	Description
Goods		
1	CG01AgMinDirection	Agriculture and mining direction
2	CG02AgMinOutput	Agriculture and mining output
3	CG03ConDirection	Construction direction
4	CG04ConOutput	Construction output
5	CG05MfgDirection	Manufacturing direction
6	CG06MfgOutput	Manufacturing output
Services		
7	CS07TCUDirection	Transport and utilities direction
8	CS08TCUOutput	Transport and utilities output
9	CS09WsOutput	Wholesale output
10	CS10RetailOutput	Retail output
11	CS11FIREOutput	Finance insurance and real estate output
12	CS13OthServOutput	Other services output
13	CS14HealthOutput	Health output

No	Commodity Name	Description
14	CS15GSEdOutput	Grade school education output
15	CS16HiEdOutput	Higher education output
16	CS17GovOutput	Government
Receipts		
17	CF18TaxReceipts	Tax receipts
18	CF19GovSupReceipts	Government support funding
19	CF20InvestReceipts	Investment receipts
20	CF21ReturnInvestReceipts	Returns on investment receipts
21	CF22CapitalTransferReceipts	Capital transfers
Labor		
22	CL23WhiteCollar	White collar (SOC:11-23, 25, 43)
23	CL24Services	Services (SOC: 27,33,37,39)
24	CL25Health	Health (SOC:29,31)
25	CL26Retail	Retail and food (SOC:35,41)
26	CL27BlueCollar	Blue collar (SOC:45,47,49,51,53)
27	CL28Military	Military
Space		
28	CA29AgMin	Agricultural and mining space
29	CA30Indust	Industrial
30	CA31Retail	Retail
31	CA32Office	Office
32	CA33Instit	Institutional
33	CA34Military	Military space
34	CA35DetResid	Detached residential
35	CA36HiDenResid	Higher density residential

An overview of the AEFT developed for the ARC PECAS model is shown in Figure 1. The left half shows the "make" or production of commodities by the activities, while the right half shows the "use" or consumption of commodities by the activities. Color codes indicate particular interaction between activities and commodities for the ARC PECAS model and are described in Figure 1.

The file reporting the data and the procedure to develop the AEFT for the ARC PECAS model is called **"PecasIMPLANtemplt_201204. xls".**

Figure 1. Overview of the AEFT for the ARC PECAS Model



3. Overview of procedures to develop relationships between households, labor categories and space categories

In the PECAS framework, households consume (live in) residential space (housing) and provide labor within different occupations. These employees are required by the industries, where they work in non-residential space consumed by their employer. These relationships were developed for the ARC PECAS model using procedures described in this section.

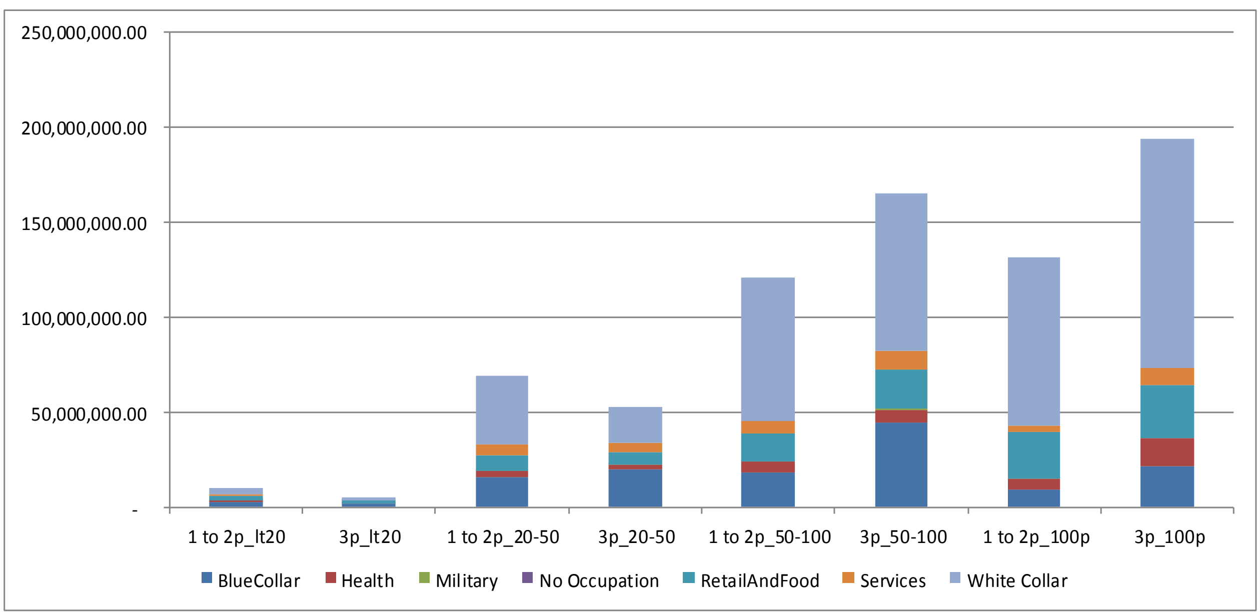
Households providing labor

The US Census PUMS (Public Use Microdata Sample) data for the Atlanta Region contains the required data regarding households (size, income, occupations and wages) and housing (type, age, rooms, bedrooms and other data). These data were processed in order to define the household categorization (by size and income) and the corresponding number of households in each category. The procedures involved in the processing included: identifying the income band and the size band for the households and adding the corresponding wages by occupation for each one, allowing the calculation of two specific relationships:

- Total wages by household size and by income and occupation
- Disaggregate labor production by households into occupation categories, which is called in the "labor make table" (Figure 3) in PECAS jargon

The processing of the PUMS data is reported in a SQL script file called **"PumsProcessing.sql"**.

Figure 2. Average wages by household category and occupation



The labor make table indicates the proportion of employment by occupation being provided by each household category. These proportions were applied to the IMPLAN total earnings data to calculate the employment numbers being produced by household categories in each occupation category in the ARC PECAS model for the base year.

PUMS data do not report residential unit size in square feet, and they only contain 5% of the households by PUMA (the geographic area associated with the PUMS data). These two issues were addressed in the next subsections by performing estimations to predict housing size and developing the synthetic population for the households in the Atlanta Region.

Developing a sample: households using housing types

This process also uses US Census PUMS data but with the purpose of calculating the number of households living in specific building types. The categorization of building types used in the synthesis of households and their space use is shown in Table 3. (Note that the ARC PECAS Model does not yet use this many categories for residential space.)

Table 3. Building type for the Atlanta Region

Housing type category	Housing type description
detached-A	Detached units
attached-A	Attached units
unit2-A	Building with 2 units
unit3to4-A	Building with 3 to 4 units
unit5to9-A	Building with 5 to 9 units
unit10to19-A	Building with 10 to 19 units
unit20to49-A	Building with 20 to 49 units
unit50ormore-A	Building with 50 units and more
mob_h-A	Mobile home
boat_rv_van-A	Boat, recreational vehicle or van

The processing performed to categorize households based on income, size and building type is reported in the script file called **"build_samples.sql"**. The resulting view (table) is called **`hh_income_cat`** in the schema called *pums*.

Developing a sample: households using residential space

The housing size estimation was performed using PUMS (Public Use Microdata Sample) data in combination with coefficients from multiple regressions to estimate unit size by unit type.

The coefficients were estimated to predict housing size by building type (unit type) by SANDAG using the national American Household Survey data. The independent variables are the following:

- Household income
- Number of rooms
- Number of bedrooms
- Type of building
- Dwelling age

The coefficients and the equations employed for each housing type are reported in the script file called **"pums_hh_sqft.sql"**.

The same variables were used in the Atlanta Region, using PUMS data to estimate the unit size in square feet (residential square feet) of the household sample data for the ARC PECAS model. This sample was used to generate the synthetic population of the households in the region.

Synthetic population of the households in the Atlanta Region

Population synthesis is a process of generating a representation of a complete and disaggregate population by combining a sample of disaggregate members of a population. This sample can include any type of variables: persons, dwellings, families or others, combining them so as to match key distributions for the entire population. This tool is normally applied in land use modelling, transportation modelling and in similar contexts.

The population synthesizer program used for the ARC PECAS model is the one developed by HBA Specto Incorporated, which uses the combinatorial optimization approach (John E. Abraham, Stefan, and Hunt 2012). This has been applied to synthesize population and/or employment for several studies in different contexts and it has been useful in achieving workable datasets. Some of the studies where it has been applied are: the California Statewide Travel Demand Model (CSTDM), for Oregon State in the context of the Transportation and Land Use Model Improvement Program (TLUMIP), and for the City of Calgary.

The process requires a sample of the population to expand, and targets to match by geographies. Different geographies can be used for different targets. For this case the PUMS household sample was used for its data on household income and size, and the type of building each household lives in – building type and number of rooms.

A trial population was created from the disaggregate PUMS data, and the overall goodness of fit was measured across all marginal targets. A summary table showing the targets, as well as the variable, geography type and data source of each target is presented in Table 4.

Table 4. Targets for the synthetic population of households for the ARC PECAS model

Variable	Targets	Geography type	Source /Date
Household totals	Number of households by income and size	TAZ	ARC Travel Model
Dwellings	Number of vacant units Number of units	TAZ	ARC Travel Model
Households by income and size	Income 1, 1 people Income 1, 2 people Income 1, 3 people Income 1, 4 people Income 1, 5 people Income 1, 6 and plus people (These 6 targets are present for each of the 4 income categories. There are 24 targets associated with these target amounts)	TAZ	ARC Travel Model
Proportion of units by type where total households were living	detached-A attached-A unit2-A unit3to4-A unit5to9-A unit10to19-A unit20to49-A unit50ormore-A mob_h-A boat_rv_van-A	Census Track	2000 Decennial Census
Proportion of units by type where households by type were living	detached_29-A attached_29-A unit2_29-A unit3to4_29-A unit5to9_29-A unit10to19_29-A unit20to49_29-A unit50ormore_29-A mob_h_29-A boat_rv_van_29-A (These 10 targets are present for each of the 8 household types: 29 to 36. There are 80 targets related with these proportions)	PUMA	PUMS (US Census)
Proportions of the number of dwellings with a specific number of rooms	rooms1-A rooms2-A rooms3-A rooms4-A rooms5-A rooms6-A rooms7-A rooms8-A rooms9p-A	Census Track	2000 Decennial Census

The spread sheet containing the processing data for the targets employed to synthesize the population of households for the ARC PECAS model is called **"To build targets with Entropy.xlsx".**

Units from the trial population are swapped with units chosen from the disaggregate sample, and if the measure of fit improves, the swap is made. Iterations continue until an acceptable fit to the targets is established and the entropy of the solution is maximized. This particular technique has a variety of features that make it useful and suitable in working with different data sources associated with multiple geography levels (John E. Abraham, Stefan, and Hunt 2012).

The resulting synthetic population, with its estimate of dwelling sizes for each household, was aggregated:

- by TAZ, to provide model input data on the amount of space by each PECAS residential space type in each TAZ. (Thus, the TAZ estimates of space reflect TAZ level observed variables, as well as Census Tract and PUMA information.)
- by household category and PECAS space category, to provide information to the Aggregate Economic Flows Table on the use of space by households.

Industry using non-residential space as locations for employees

The employed people in PUMS households are classified according to the industry of their employer, as well as their own occupation. This information is used to understand how employment activity requires different types of space. Certain employment categories (e.g. agriculture, industrial, office, retail, institutional and military) require production space to undertake their primary activity, yet also use office space for their employees in "white collar" occupations. Employment data available in the USA reports employees in production industries in downtown office locations, where only office space exists, and true production of the industry's product is not possible. In PECAS, industries that require substantial quantities of both production space and office space in different locations are split into two portions, the *production* component and the *management* component. The allocation of the two components to zones is described in this section.

Calibrating the exponent to split labor based on the space type used

One of the typical procedures required in a PECAS model is to split the labor depending on the type of space where the job takes place. The objective is to separate the employees that work in office space from the ones working in non-office space in each industry of the Atlanta Region using a labor split exponent LE.

The labor split exponent was used to calculate the employment by TAZ for each industry working in non-office space $Emp\ Non\_office_{\ i,z}$, taking the product of the total employment by industry and by TAZ and the office split factor to the power of the labor split exponent LE using the following formula:

$$Emp\ Non\_office_{\ i,z} = Emp_{\ i,z} * Office\ Split\ factor^{\ LE}$$

The office split factor was calculated in a separate procedure using LandPro data for all the counties in the Atlanta Region. The idea was to define relationships between office space and other non-residential space types in each TAZ. Four split factors were calculated dividing the columns presented in Table 5. The most suitable of these factors was assigned to each industry category in the process to calculate the employment working in non-office space.

Table 5. Office split factors to calculate the employment working in office space

Office split factor	Numerator (a)	Denominator (b)
Industrial versus office space	Industrial+**0.5***Industrial_Commercial+**0.1***(TCU+Transitional+Urban_other)	Industrial + Industrial_Commercial +Commercial+Urban_other+Institutional_Intensive+**0.1***Institutional_Extensive+**0.1***SUM(TCU+Transitional+Urban_other)
Agriculture and mining versus office	Agriculture+Quarries	Agriculture + Quarries + Commercial + Industrial_Commercial*2+ Institutional_Intensive
Institutional versus office	Institutional_Intensive +**0.1*** Institutional_Extensive	Institutional_Intensive +**0.1*** Institutional_Extensive + Commercial +**0.5*** Industrial_Commercial

Once the employment in non-office space was calculated, the employment working in office space was calculated as the difference between the total employment by TAZ and the employment working in non-office space.

The labor split exponent LE to split the labor into the referred space categorization was calculated by iterations using an optimization process. This involved each industry split into the office space and the non-office space. The optimization process used the solver analysis tool in Excel. The labor split exponent was adjusted by iterations, until the total (model-wide) division of labor between office and non-office matched the total division of the industry's employees between white collar occupations and blue collar occupations from the PUMS data.

Some post processing was performed afterwards. The details are in the spreadsheet called **"EmploymentAllocationV1B-noSpaceInputs.xlsx"** (Note the spreadsheet filename is misleading – the process did use inputs of space type by TAZ from LandPRO, as discussed above.)

Calibrating average space use rates for non-residential space

Space use rates were calibrated for the non-residential space categories such as agriculture, industrial, office, retail, and institutional. An optimization process was applied where the objective was to minimize the error between the estimated space and the observed space by TAZ, using observations from the Fulton County with these five space types by TAZ.

The estimated space by type and by zone was calculated by the five categories by multiplying the employment (by industry) and by TAZ by the space use rates. Four different space use rates were then estimated, to minimize the sum-of-squared error between the quantity of space-by-type observed in the Fulton county TAZs and the estimated space. The employment used in this procedure was already split into production and management, in order to differentiate office space from production space. The resulting space use rates for the ARC PECAS model are shown in Table 6.

Table 6. The calibrated space use rates for the Non-residential space

Space	Space Use rate (Sq Ft / employee)
Office	497
Industrial	709
Retail	804
Institutional	400

The procedure is reported in the same spreadsheet as the process to calculate the labor split exponent, called **"EmploymentAllocationV1B-noSpaceInputs.xlsx".**

Note that the resulting amounts of space by zone were used as interim inputs and calibration targets not as inputs to the final model (see section 8).

4. Overview of procedures to develop space rents as calibration targets

Two different procedures were developed to calculate average rents per space type and Super District, one for the residential space and one for the non-residential space. The PECAS AA module requires rent information by zone that is an indication of the attractiveness of the zone, representing the spatial interaction possibilities of the zone (the advantages of the location in terms of selling or purchasing Commodities from other activities in other locations) and intrinsic characteristics of the environment of the zone itself. Thus, the rent information used in the AA model must control for other variables that relate more to the type and quantity of development than the location and environment.

(The estimated coefficients associated with the control variables are used in the SD module; SD contains a full microsimulation of parcels and their improvements, so specific types of dwellings can be more or less desirable in SD).

Rent estimation for residential space

For residential space, the rent estimation controls for the following housing variables:

- Age of buildings (older buildings generally less desirable)
- Size of residential buildings (larger houses are more expensive and zones with many larger houses will have higher average rents than equally attractive zones containing many smaller houses)

- Type of dwelling (single family homes tend to be more expensive than similary sized multifamily homes, due to the utility value or investment value of associated gardens)

The residential rent estimation uses the following data for Census blocks from Census SF3:

- Number of units by 9 dwelling categories, based on the number of rooms of each unit. Eg. Dwellings with 1 room, dwellings with 2 rooms, and so on until dwellings with 9 rooms or more.
- Aggregate monthly rent per block
- Number of dwellings being single family detached (SFD)
- Number of dwellings being other category different from SFD
- Median of the year built of the dwellings rented in the block

Multiple regressions were performed to calculate residential rents by census track as well as six adjustment parameters. Specifically, there is a tract-rent constant estimated for each census tract *t*, where block *b* is contained within *t,* and these rents are adjusted being multiplied by the following estimated parameters:

- Three rent adjustment factors for the block *b*, based on the number of rented dwellings in block group *b* with number of rooms *r*
- A rent adjustment factor for the age of the dwellings
- A rent adjustment factor for the portion of housing value not affected by age
- A rent adjustment factor for the structure type (SFD vs other)

A brief document reporting the mathematical equations involved in this rent estimation for the residential space is included as an appendix of this document (Appendix 1). As an example, the rents estimated for three census tracts (before adjustment) are shown in Table 7. And the parameter values to adjust these rents are shown in Table 8. A map of the adjusted rents for residential space is shown in Figure 3.

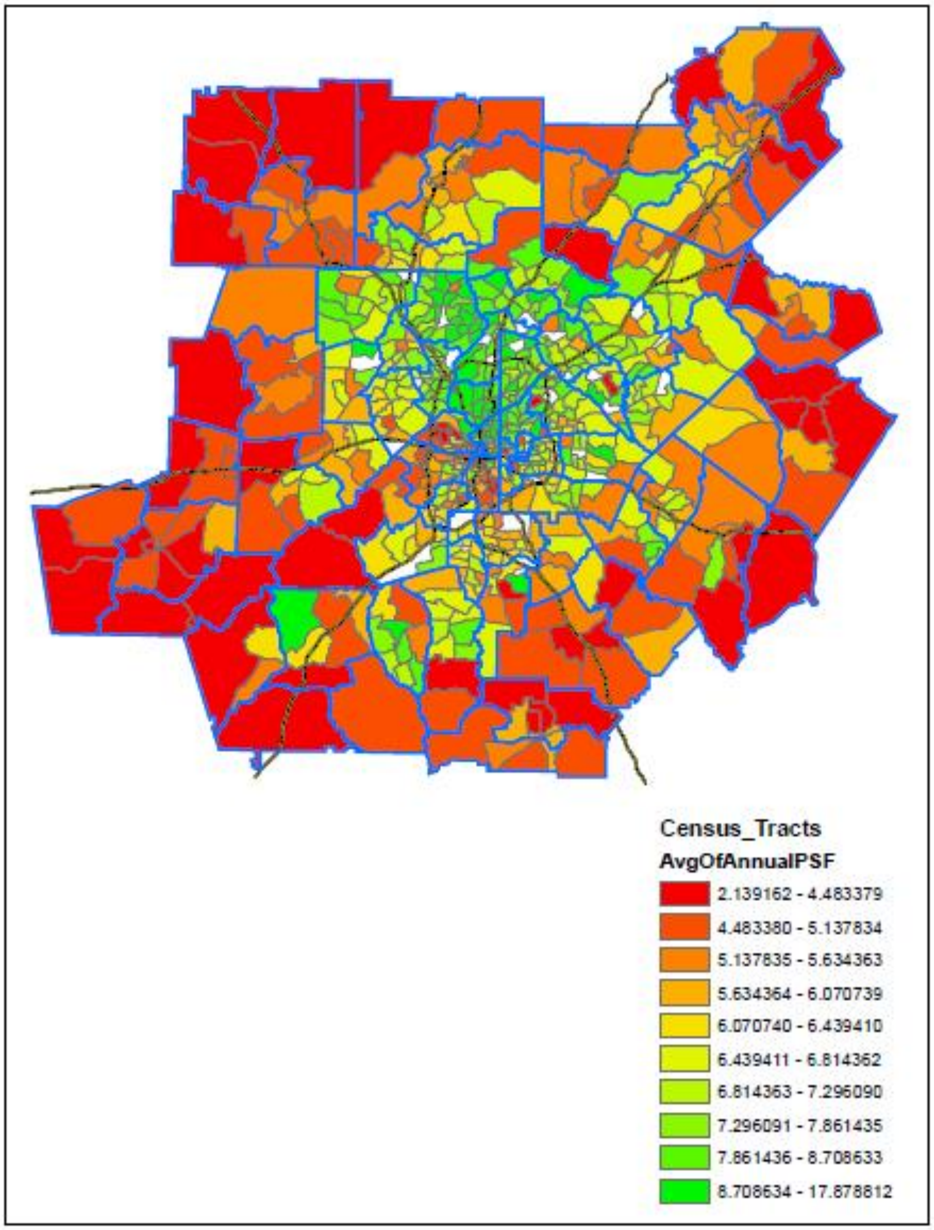
Table 7. Estimated rents by census tract

Census Track	Estimated rent
113_1401.01	729.4017
113_1401.02	810.5785
113_1402.03	772.8256

Table 8. Parameter values estimated to adjust rents by census tract

Parameter name	Estimated Parameters
size ϕ_1	0.150719
size ϕ_2	0.001272
size ϕ_3	2.85E-09
ageAdj	0.012353
ageIndep	0.181928
rent_allother	1.28678

Figure 3. Average price by LUZ for residential space



The files involved in the rent estimation for the residential space includes: **"PECAS_BGs_Rents_BaseData.xlsx"**, which has the input data, and **"RentRegression.py"** containing the python script to perform the rent estimation for the residential space.

Rent estimation for the non-residential space

For the non-residential space, a rent estimation procedure was developed using rent observations for buildings from Costar data. This rent estimation was performed for the following space types: retail, office and industrial.

The predicted rent is defined as as:

$$\text{Pred R}_{\text{n}} = \text{Ave R}_{\text{z}} \text{ x AgeMod x ClassMod}$$

$$\text{Age Mod}_{\text{n}} = \begin{cases} \text{P2} + (1 - \text{P2}) \cdot (1 - \text{p3})^{\text{age}}, \text{if age is valid} \\ \text{P1, if age is missing} \end{cases}$$

Where,

Pred Rn	= Predicted rent ($/ Sq Ft)
Ave R z	= Average rent for superdistrict
Age Mod	= Age modifier term
Age	= Effective age of building
P1	= Effective age is missing
P2	= parameter associated with the part of the equation independent from age
P3	= parameter associated to the decay in rent due to age
n	= number of observations (there were several by zone)
ClassMod	= Estimated modifier for the each “class” of building, used for office space

An optimization procedure was applied to minimize the error between the predicted rents and the observed rents, while adjusting the parameters associated with the rent modifier terms. The total error is squared and weighted by the square of the rentable building area. The sum across all of the observations, has the following form:

$$\sum_{n} E\ (SSE) = \ [Obs\ R - (Pred\ R)]^2 \quad x \quad RBA^2$$

Where,

SSE	= Summation of the squared error
Obs R	= Observed rent
RBA	= rentable building area

Several iterations were performed until the SSE is minimized. The final parameters calculated for each space type (Table 9).

Once the parameters for the rent modifier by observation due to the building age were estimated, a parameter for the adjustment at average age by super district or land use zone (LUZ) was calculated as:

$$\text{P4}_{\text{z}} = (\text{P2} + (1 - \text{P2}) \text{ x } (1 - \text{p3})^{\text{Ave Age}})$$

Where,

P4	= parameter for the rent adjustment at average age by LUZ
Ave Age	= Average building age of the observations

The estimated parameters for the rent adjustment at average age by LUZ for the 3 space types are presented in Table 9. The rent is also adjusted by space type based on the class factors (CF) indicated in Table 10.

Table 9. Estimated parameters for the rent estimation by space type

Space type	Retail	Office	Industrial
Number of observations	1470	2053	1288
Parameters by observations			
Parameter associated with missing age data (P1)	0.524427984	0.906689	0.686468
Parameter associated with the part of the equation independent from age (P2)	0.539571106	0.012612	0.847639
Parameter associated to the decay in rent due to age (P3)	0.099787243	0.004687	0.291035
Average Building Age	19.56872852	20.63775	23.70331
Rent adjustment at average age by LUZ (P4)	0.598422324	0.908752	0.847683

Table 10. Class Factors for industrial, flexible and office space

Space	Rent Modifier
Industrial (Baseline)	1.000000000
Flex (Industrial –Retail)	1.552717025
Office class A	1.067200442
Office class B (Baseline)	1.000000000
Office class C	0.804144066

The rent estimation by LUZ for the Atlanta region was performed multiplying the average rent by LUZ by the parameter for the rent adjustment at average age by LUZ.

To enable a simple parameter search in Excel over the parameters of the age and class modifiers, the system of equations was solved algebraically for the zonal average adjusted rents of new property *Ave Rz*, and the zonal average adjusted rents for properties of an average age *Rz.*

$$R_Z = Ave\ R_Z\ .\ P4\ .\ CF$$

$$Ave\ R_Z = \frac{Agg\ R_Z}{W\ R\ Adj}\ .\ CF$$

$$Agg\ R_Z = \sum_n Ave\ R_n\ \ x\ Age\ Mod\ x\ RBA^2$$

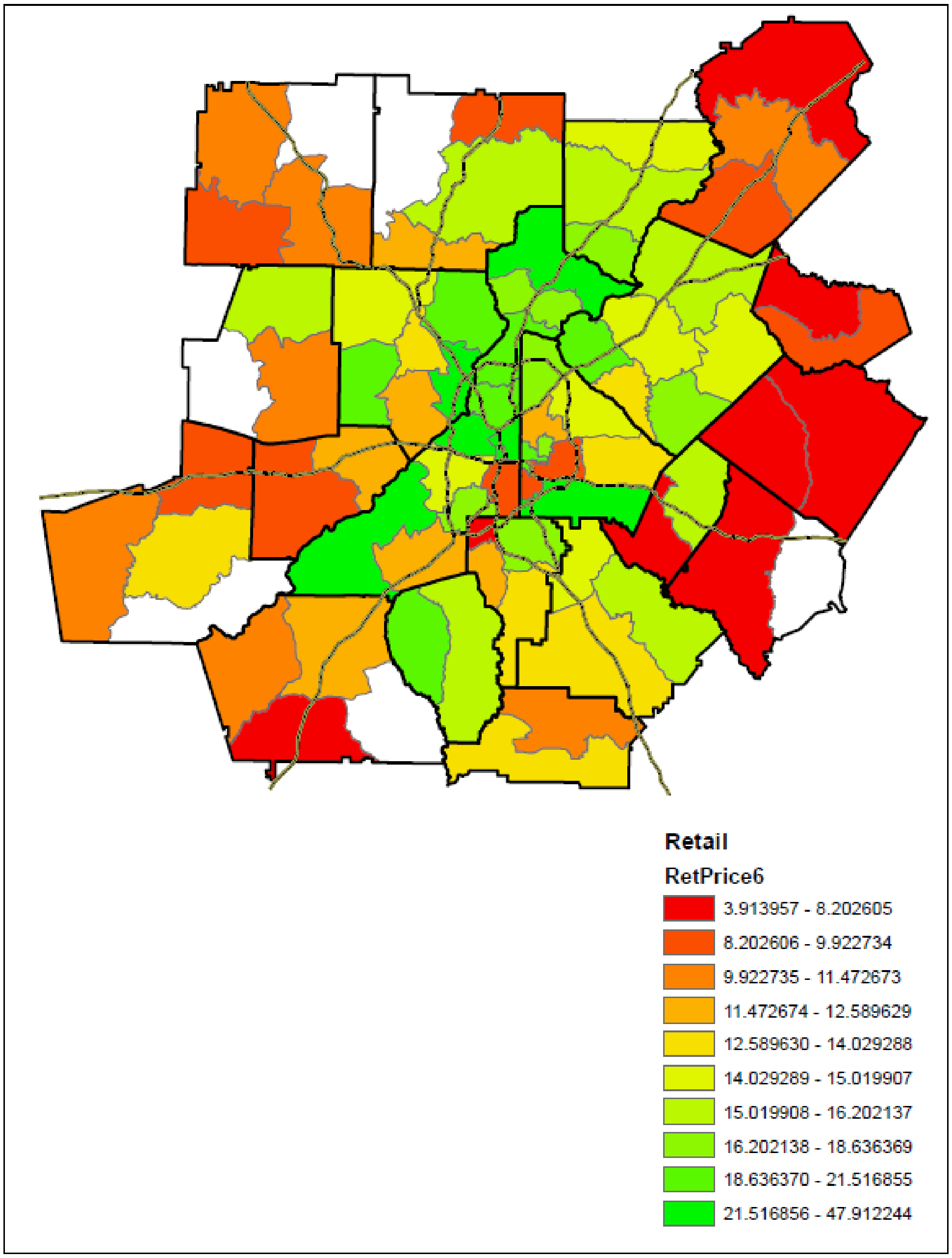
$$W\ R\ Adj\ _Z = \sum_n (Age\ Mod_n\ \ x\ \ RBA_n)^2$$

Where,

R z = Rent by LUZ
Ave R z = average rent by LUZ
Agg Rz = aggregate rent by LUZ,
W RAdj = Weighted Rent Adjuster
CF = Class factor

The derivation of the formula is shown in the Appendix 1. The final rent by super district for typically aged buildings, Rz, is reported respectively in Figure 4 for retail space, Figure 5 for office space and Figure 6 for industrial space.

Figure 4. Average price by LUZ for retail space



The details of the procedures performed for the rent estimation for the non-residential space are contained in the spread sheet called **"NonResidentialRentEstimation .xls"**.

Figure 5. Average price by LUZ for office space

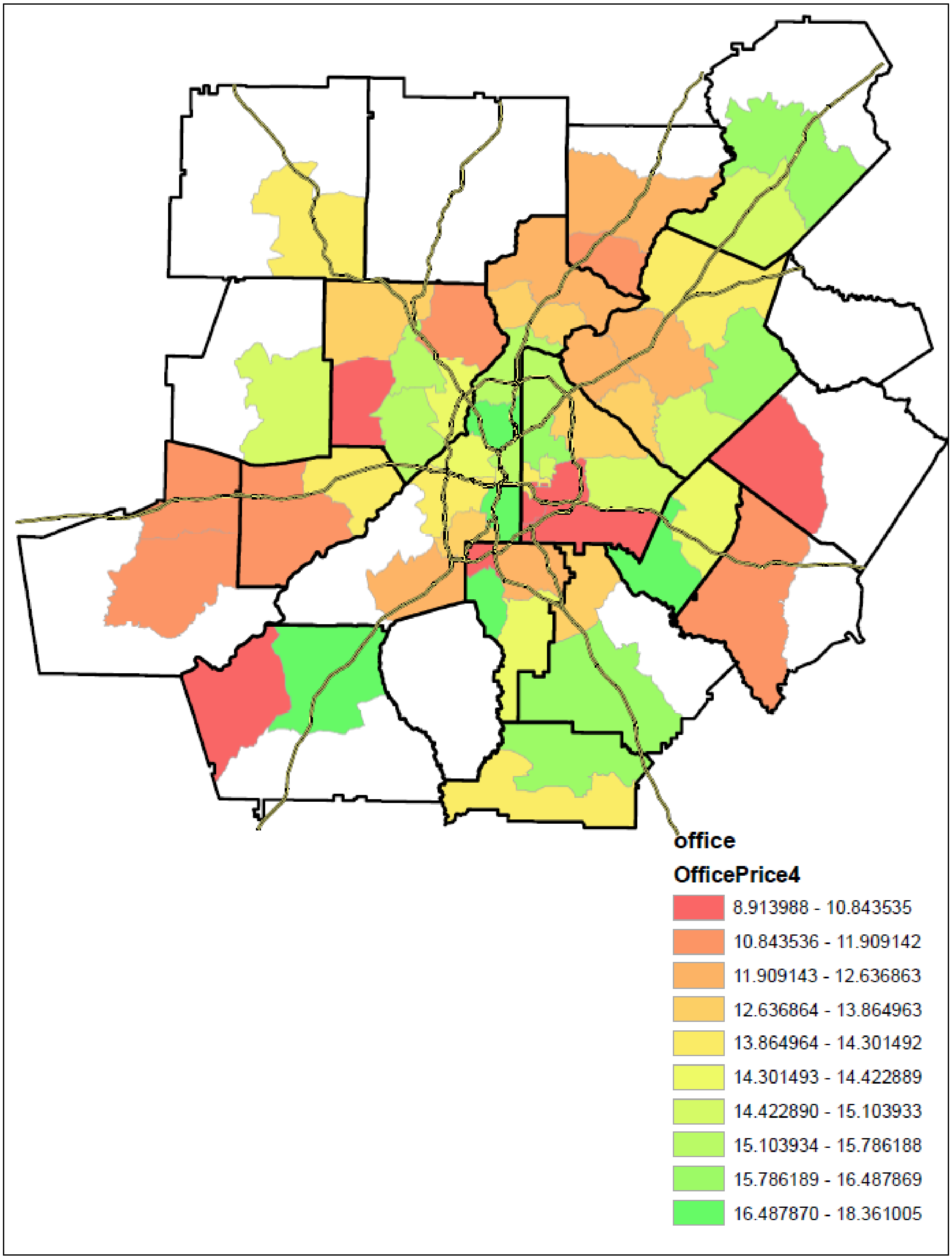
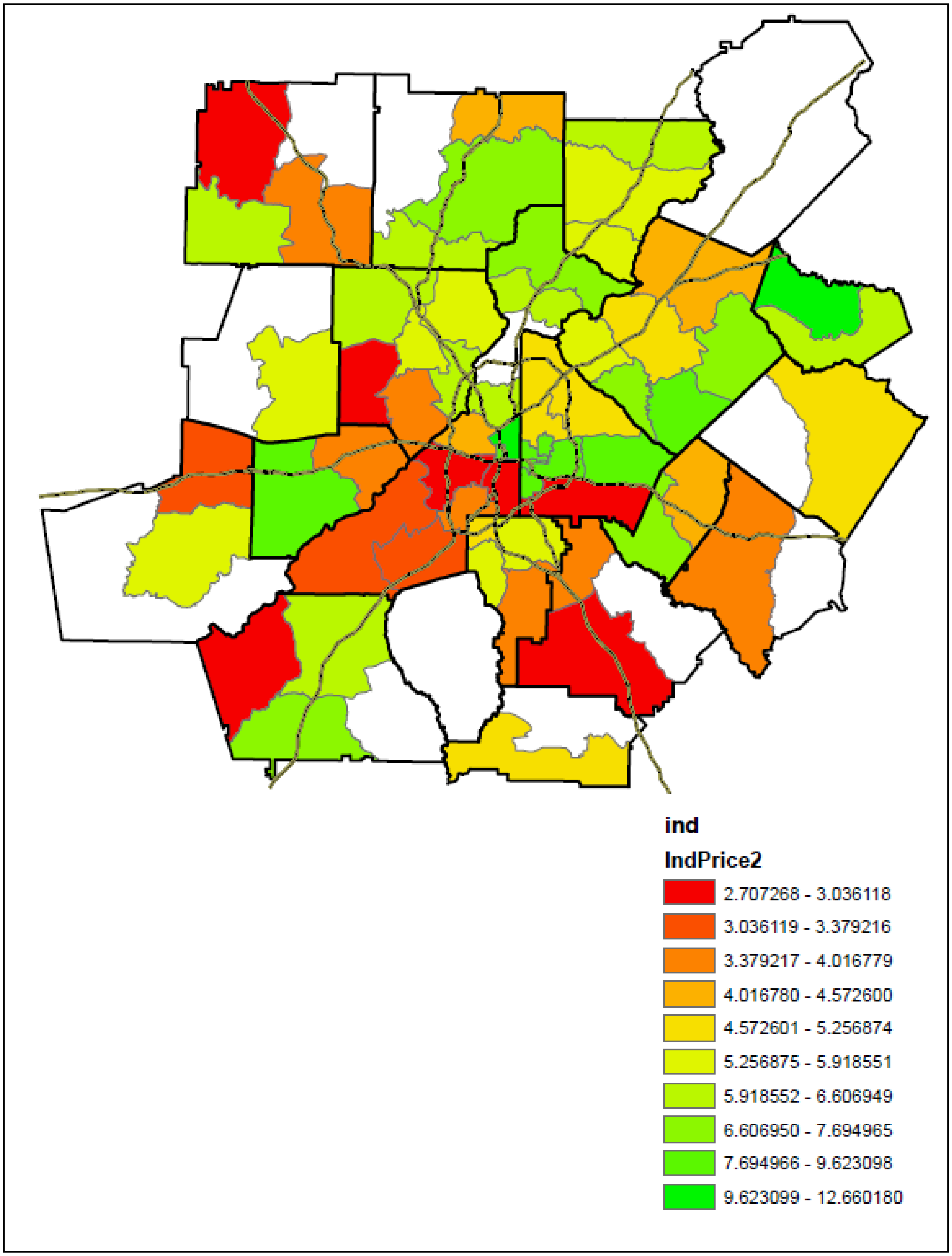


Figure 6. Average price by LUZ for industrial space



5. Overview of the Transport Utility Functions

The activity allocation (AA) module requires values for the utility of transporting a unit of each commodity category considered in the model. These utilities are calculated for each zone to zone interaction using transport cost coefficients and skim interchange values determined by the transport model for Atlanta.

There are 3 types of coefficients involved in the utility for buying or selling a unit of commodity c in an exchange location k produced in a zone z. These coefficients are: the size coefficient, the price coefficient, and the transport coefficient.

The coefficients for size are often set to 1 for both the buying and the selling coefficients as it was done for Atlanta (a zone twice as big should get twice the share, all other things being equal), while the coefficients for price are set to -1 for the buying coefficient and +1 for the selling coefficient, assuming a dollar of expenditure/earning is as important as a dollar of expenditure on transport, and indicating the correct sign on the price when an amount of commodity is bought or sold. The coefficients for transport are different from zero and need to be estimated. Some financial commodities were assumed to move with no cost, the other cost functions were estimated for the Atlanta Region.

The procedure to estimate the transport coefficients is described in detailed in a document called *Task 11. Prepare Transport Cost Coefficients*. For each commodity, a series of skim values - such as monetary cost (toll, parking charges, transit fare, and others), travel time, travel distances, and composite utilities (also called mode choice logsums) were multiplied by transport cost coefficients and combined in order to establish the utilities for transporting commodities considered in the AA module.

The general form of the transport utility equation is:

$$Tran_{c,j,k} = \kappa 1_c \cdot IntAtt1_{j,k} + \kappa 2_c \cdot IntAtt2_{j,k} + \kappa 3_c \cdot IntAtt3_{j,k} + \kappa n_c \cdot IntAttn_{j,k}$$

Where:

$Tran_{c,j,k}$	= the utility for transporting a unit of commodity c from zone j to zone k;
$IntAtt1_{j,k}$	=value for interchange attribute 1 from zone j to zone k;
$IntAtt2_{j,k}$	=value for attribute 2 from zone j to zone k;
$IntAtt3_{j,k}$	= value for attribute 3 from zone j to zone k;
$IntAttn_{j,k}$	= value for attribute n from zone j to zone k;
$\kappa 1_c$	= utility function coefficient for the sensitivity to attribute 1 when transporting a unit of commodity c;
$\kappa 2_c$	= utility function coefficient for the sensitivity to attribute 2 when transporting a unit of commodity c;
$\kappa 3_c$	= utility function coefficient for the sensitivity to attribute 3 when transporting a unit of commodity c;
κn_c	= utility function coefficient for the sensitivity to attribute n when transporting a unit of commodity c;

All of the attributes used to calculate the utility for transporting a unit of commodity c were provided as a network skim values from the transport model.

In general, transport acts as an impedance, so utility values for $Tran_{c,j,k}$ are likely to be negative. Time, distance, and monetary costs as interchange attributes are all positive or zero. Increases in these attribute values make transport more unattractive, so the transport coefficients associated with these attributes are also negative. Mode choice logsums as interchange coefficients decrease as travel become less appealing, so the associated transport coefficients are positive. (Zero is an arbitrary point in mode choice logsums, but in the ARC model, as in most models, the zero point is set to be the utility of a zero distance trip by a single mode, and trips of non-trivial distance have negative mode choice logsums.)

The procedure to calculate the transport cost coefficients depend on the type of commodity. In general, there are 4 types of commodities: Goods, Household obtained services, labour and worker delivered services. For the ARC PECAS model the 22 commodities were organized in these four types and the inputs used for the transport cost coefficient estimation are presented in Table 11.

Table 11. Inputs for the transport coefficients for each commodity type

Commodity Type	Commodity Name	Transport Coefficient Inputs
Goods:	CG02AgMinOutput CG04ConOutput CG06MfgOutput CS08TCUOutput CS09WsOutput	Cost per Mile ($/Mile) per type of truck Cost per hour ($/Hr) per type of truck Cost per weight ($/ton-Mile) per commodity Average use of truck capacity by commodity and by type (capacity) Distribution of trucks by commodity and by capacity Mean hourly wage rate for heavy and light trucks
Household obtained services:	CG10RetailOutput CS13OtherServOutput CS14HealthOutput CS15GSEdOutput CS16HiEdOutput	Total annual consumption (money value) of commodities by households Number of annual visits Transport Money Cost Coefficient explicit in the mode choice model
Labour:	CL23WhiteCollar CL24Services CL25Health CL26Retail CL27BlueCollar CL28Military	Distribution of annual wages paid to occupations by income and vehicle owning Number of annual visits by income and vehicle ownership Transport Money Cost Coefficient explicit in the mode choice model
Worker delivered services:	CG01AgMinDirection CG03ConDirection CG05MfgDirection CS07TCUDirection CS11FIREOutput CS17GovOutput	Annual production (money value) of commodities Number of annual visits Transport Money Cost Coefficient explicit in the mode choice model

In the ARC PECAS model these transport coefficients were originally calculated using data and skims generated by the 4-step transport model developed for the region. For this model a zone system with around 2000 zones was developed. The first transport coefficient estimation performed for the ARC PECAS model is contained in the file called **"ARC PECAS.TCC.2.xls".** The current version of the ARC PECAS model is calibrated using this 4-step transport model.

More recently ARC developed a new transport model (ABM) with a more detailed zone system including almost 6,000 zones. Since the ARC PECAS model was already calibrated to the 4-step transport model, the goal was to develop more detailed transport utility functions using the data from the new transport model to predict the skims from the previous model, to use the new ABM-generated skims without performing a full recalibration of the ARC PECAS model.

In order to predict the old skims multiple linear regressions were estimated with the new skims. Python code was written to estimate and search the optimum model using skims from both transport models. The new model produces 4 different skims differentiating the skims by vehicle ownership and income. The following criteria were used to include the new transport skims in the transport utility function for the commodities by type (Table 11) in the ARC PECAS model:

- The good commodities keep the same previous three terms in the function, one for the distance, one for the time and one for the toll (Table 13)
- The worker delivery services (management part of the goods) commodities include the skim for the high income with auto and a constant (Table 13)
- The household obtained services commodities uses a constant and two skims, the no autos low income and the one with auto and high income (Table 13)
- The labour commodities use the four skims since the proportion of employment in each category was known (Table 12).

The adjustments of the transport utility functions were performed by parts. For the goods, the transport coefficients were kept and only the skims were changed; the description of the adjustments made for the transport coefficients used for the worker delivery service commodities, for the household obtained services and for the labour commodities is presented in the Appendix 3 of this document. The transport utility for the labour commodities were adjusted using the proportion of labour in each category (Table 12) and the estimated regression coefficients reported in the Appendix 3.

Table 12. Proportion of labor by income category

Labor Catgeory	lt20	20-50	50-100	100p	Total
BlueCollar	3.48%	26.62%	47.15%	22.76%	100.00%
Health	1.93%	13.51%	31.65%	52.91%	100.00%
Military	0.57%	25.24%	56.26%	17.93%	100.00%
RetailAndFood	2.91%	14.76%	33.22%	49.11%	100.00%
Services	4.34%	25.04%	41.45%	29.17%	100.00%
White Collar	1.03%	12.90%	37.02%	49.05%	100.00%
Grand Total	1.96%	16.29%	38.23%	43.51%	100.00%

The current transport utility functions for the commodities in the ARC PECAS model use the ABM skims and the estimated coefficients reported in Table 13.

Table 13. Updated skims from the ABM model

Commodity	Skim_1	TC1	Skim_2	TC2	Skim_3	TC3	Skim_4	TC4	Skim_5	TC5
CG01AgMinDirection	one	0.0115148	OP_VHI_A	0.068533538		0		0		0
CG02AgMinOutput	AMHWDIST	-0.000822519	AMCOMPTIME	-0.000133071	AMTOLL	-0.000440632		0		0
CG03ConDirection	one	0.011514827	OP_VHI_A	0.068533538		0		0		0
CG04ConOutput	AMHWDIST	-0.000055784	AMCOMPTIME	-0.000010084	AMTOLL	-0.000033391		0		0
CG05MfgDirection	one	0.011514827	OP_VHI_A	0.068533538		0		0		0
CG06MfgOutput	AMHWDIST	-0.000074849	AMCOMPTIME	-0.000012835	AMTOLL	-0.000053036		0		0
CS07TCUDirection	one	0.011514827	OP_VHI_A	0.068533538		0		0		0
CS08TCUOutput	AMHWDIST	-0.000149724	AMCOMPTIME	-0.000025389	AMTOLL	-0.000095989		0		0
CS09WsOutput	AMHWDIST	-0.000014697	AMCOMPTIME	-0.000002256	AMTOLL	-0.000009936		0		0
CS10RetailOutput	one	0.022974895	OP_LI_0A	2.85398E-05	OP_HI_A	0.034435076		0		0
CS11FIREOutput	one	0.00224091	OP_LI_0A	0.000036	OP_HI_A	0.01357363		0		0
CS13OthServOutput	one	0.006318093	OP_LI_0A	7.84844E-06	OP_HI_A	0.009469642		0		0
CS14HealthOutput	one	0.001914577	OP_LI_0A	2.37832E-06	OP_HI_A	0.002869594		0		0
CS15GSEdOutput	one	0.015954779	OP_LI_0A	1.98193E-05	OP_HI_A	0.023913233		0		0
CS16HiEdOutput	one	0.011966111	OP_LI_0A	1.48645E-05	OP_HI_A	0.017934965		0		0

CS17GovOutput	one	0.016806833	OP_LI_0A	0.000270334	OP_HI_A	0.101802264		0		0
CL23WhiteCollar	one	0.014259621	P_LI_A	0.000416335	P_MI_A	0.005412356	P_HI_A	0.015404397	P_VHI_A	0.020400417
CL24Services	one	0.031129998	P_LI_A	0.003075937	P_MI_A	0.019224608	P_HI_A	0.031528357	P_VHI_A	0.022300545
CL25Health	one	0.019720892	P_LI_A	0.001199912	P_MI_A	0.008399383	P_HI_A	0.019198591	P_VHI_A	0.031797666
CL26Retail	one	0.030124787	P_LI_A	0.002355324	P_MI_A	0.01177662	P_HI_A	0.025908565	P_VHI_A	0.038470293
CL27BlueCollar	one	0.030276444	P_LI_A	0.002184705	P_MI_A	0.019662348	P_HI_A	0.03422705	P_VHI_A	0.016749407

Note: The skim names used inside of the model are shown below. Short names were used only for presentation purposes in the table.

op_veryHighInc_a_ge_w = OP_VHI_A
p_veryHighInc_a_ge_w = P_VHI_A
op_highInc__a_ge_w = OP_HI_A
p_highInc__a_ge_w = P_HI_A
pk_medInc_autos_ge_workers = P_MI_A
pk_lowInc__a_ge_w = P_LI_A
op_lowInc_0_autos = OP_LI_0A

6. Overview of procedures to develop the parcel data for the Atlanta Region

PECAS Database Format

The PECAS Space Development Module requires a database representing the land in the region. The database includes several tables with their associated attributes as it is shown in Figure 7.

The space types are defined in the ***space_types_i*** table, while the land is defined in the ***parcels*** table. The spatial policy inputs are provided in three tables that are keyed by parcel and year, representing construction costs (***parcel_cost_xref***), development fees (***parcel_fee_xref***) and zoning (***parcel_zoning_xref***). Another spatial policy input is by TAZ zone (***space_taz_limits***), this was developed in 2015 and is described in the report titled **The ARC PECAS Model Improvements, current use and ABM integration.**

This section explains how the parcel and policy tables were populated in the Atlanta SD module's database from source data.

County Parcel Assessor Data for Parcel Physical Attributes

Each of the twenty counties provided a GIS layer for their parcels. Each of the layers were imported into a spatial GIS and cleaned. The cleaning process first involved checking that parcel polygons were well defined, and then matching the polygons to the county tax assessor data. This is a many-to-many relationship: assessors sometimes have multiple records for one parcel, other times multiple parcels are assessed together. Much of the cleaning involved aggregating and splitting parcel data as appropriate to generate a single best assessment for each physical parcel.

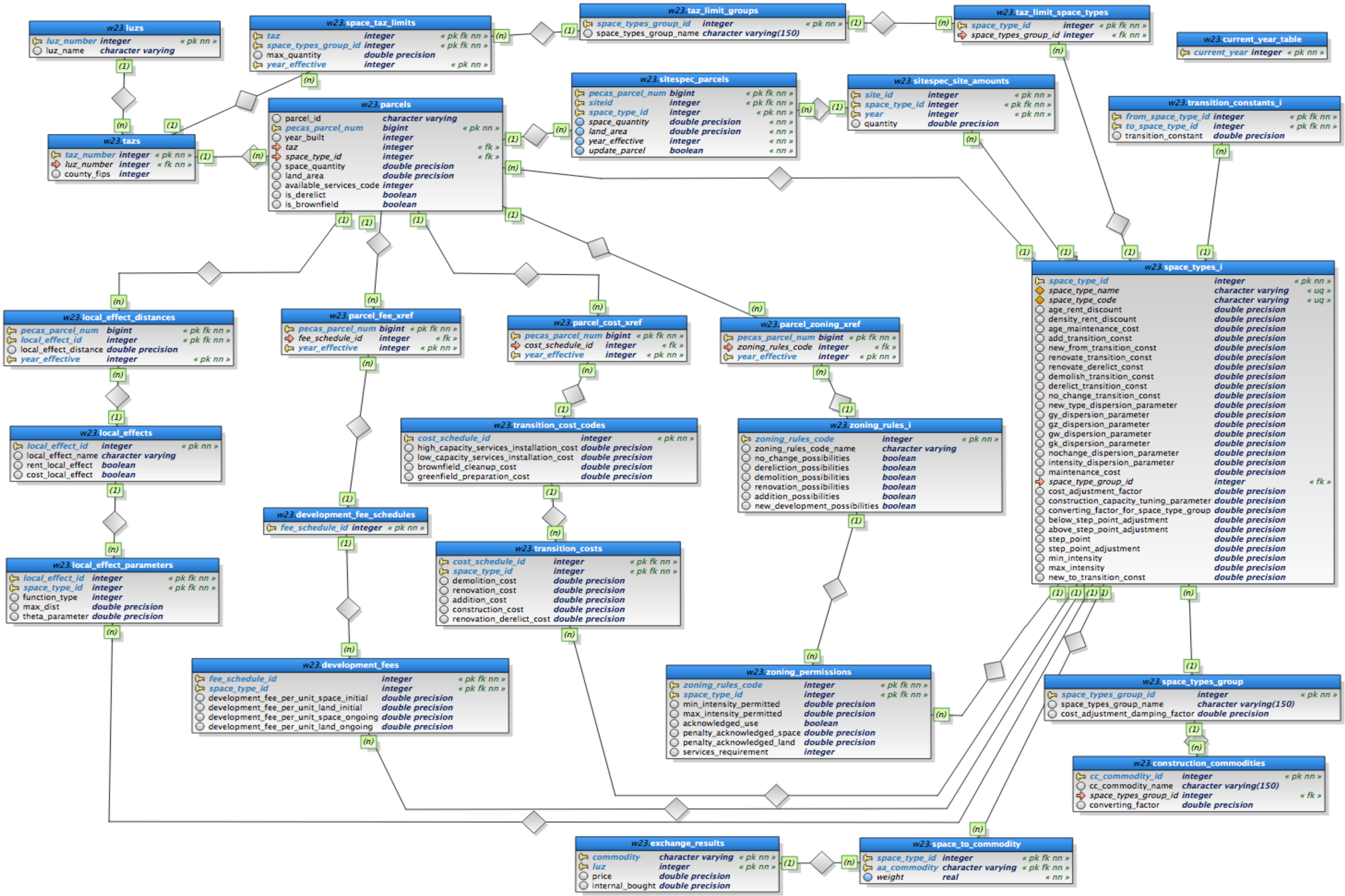
The cleaning process is an ongoing responsibility of ARC; at different times in model development there are opportunities to update the ARC PECAS model with more recently cleaned parcel data. The PECAS SD database was originally built with the 2005 parcels, and these were then updated to 2010 parcels. Clayton, de Kalb and Fulton counties have been updated with even more recent cleaned parcel data.

PECAS Parcels

From the cleaned parcels we recorded the initial PECAS space type for each parcel, the amount of space (sq ft) on each parcel, the year built of each parcel, and the land area of the parcel. PECAS requires a unique ID called the `pecas_parcel_num` for each parcel. Parcels were numbered sequentially within each county, and then prepended with the county fips code as defined by the census.

A spatial join was used to determine which TAZ each parcel is associated with, using the centroid of the parcel. This was done twice, first for the "2K" TAZ system in the old model, then for the new "mtaz10" (a.k.a. 6K taz system). The current SD uses the 6K TAZ system.

Figure 7. Structure of the relational database for the Space Development Module



Zoning information

Each county and city provides a zoning information code for each parcel.

Initially, interns at ARC were asked to review the official documents associated with each zoning code in Fulton County. Interns described each code in terms of the primary use allowed by the zoning, and the intensity that was allowed, as well as a potential second allowed use. These were entered into the PECAS SD database tables, and the initial testing and calibration of PECAS SD was done for Fulton County alone.

Once the PECAS SD module was functioning for Fulton County, the full set of all zoning codes in all counties (1,038 codes) was reviewed. A simplified set of zoning rules was defined for the use of PECAS, defined initially by the type and overall intensity of space development that is allowed. These are shown in Table 14. Each of the 1,038 individual zoning rules was associated with one of the simplified zoning rules. These zoning codes are contained in the file called **"zoning.xlsx".**

Table 14. Zoning codes and zoning description

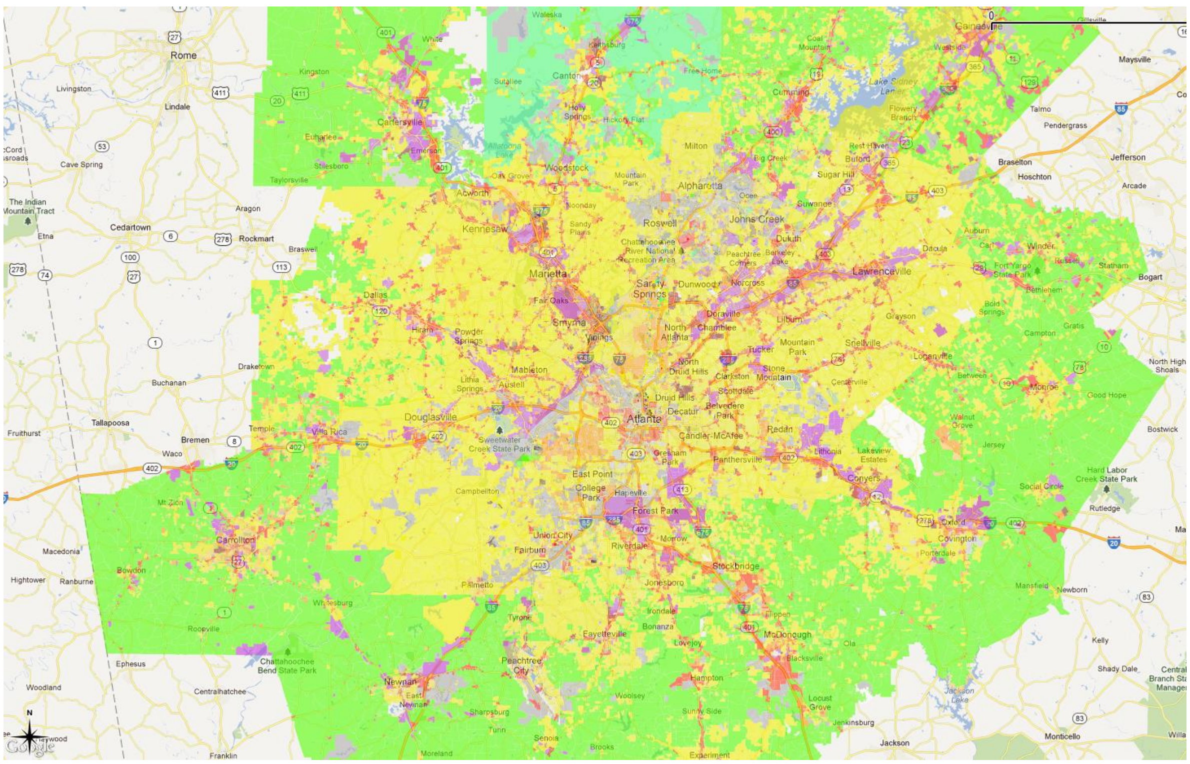
Zoning Codes	Zoning Description
10001	Conservation/protected
10002	Agriculture only (or farm residential > 40acres min lot size)
10003	Agriculture/ low density residential (acreage)
10004	Agriculture/ light industrial
10005	Low density residential (acreage)
10006	Single family residential (standard subdivision about ~4 units/acre) or low density residential (+schools)
10007	Single family residential only (standard subdivision about ~4 units/acre) (+schools)
10008	Higher density subdivision only (about 7 units/acre) (+schools)
10009	Higher or lower density subdivision (3-7 units per acre) (+schools)
10010	Low rise residential only (FAR ~0.3 to ~1.0)
10011	Low rise residential or subdivision (+schools)
10012	High rise residential only
10013	High rise residential or lower density residential
10014	Light industrial or retail
10015	"General business", light industrial or retail or office a.k.a.
10016	Low density mixed use (light industrial, retail, or residential or office up to FAR 0.7)
10017	High density mixed use (retail, residential, institutional or office up to FAR ~20?)
10018	PUD (allows almost everything, uses penalty_acknowledged_use to discourage/encourage)
10019	Light industrial (incl warehouse)
10020	Light or heavy industrial
10021	Heavy industrial only
10023	Commercial low, retail or low rise office, allows institutional
10024	Commercial high, retail or office, (low or high FAR 0.2 to 5)
10025	Office low (1-3 stories FAR 0.2 to 1.0), allows institutional
10026	Office low or high 10 stories (FAR 0.2 to 5), allows institutional
10027	High density non-residential (retail or office up to FAR ~35)
10028	Institutional/government space exclusive (includes school sites)

A selection of the most common individual zonings associated with each of the PECAS simplified zonings (outside of Fulton County) were investigated to specify quantitatively the allowed space development intensities for each allowed space type.

Within Fulton County, the 160 specific zoning categories were preserved inside of the SD database. Their correspondence to the above set of simplified categories is primarily for the purpose of presentation in maps.

A map of region level zoning is shown in Figure 8. The source file is called "**ZoningForImage.qgs**"

Figure 8. Zoning at a regional level



Zoning in future years is the same as the base year zoning in the base scenario. Alternative policy scenarios can implement modified zoning in future years as a policy test.

Local Level Effects

Under the PECAS framework rents are forecasted by the Activity Allocation module for each space type and LUZ. These prices need to be adjusted from the zonal level to the parcel level, in order to account for parcel specific and local level proximity influences on parcel rents. The rent equation used for the final calculation of the rent in each parcel is:

$Rent_h = Price_{h,z} \cdot \pi_{g \varepsilon G} LEFac_{g,h}$

Where,

$Rent_h$ = rent for updated space type h on parcel; in units of money per unit of area for space type h per year;
$Price_{h,z}$= price for updated space type h determined in AA Module for current year; in units of money per unit of area for space type h per year;
g = index of local-level effects on rent
G = set of all local-level effects on rent considered
π = the series multiplication operator, it indicates a series of terms are multiplied together in the same way that the series addition operator, Σ, indicates a series of terms are added together
$LEFac_{g,h}$ = factor adjusting proportional change in rent for space type h as a function of values on dimension relevant for local-level effect g

For the ARC PECAS model six types of local effects were defined. These effects can increase or decrease the rents depending on the space type and on the distance to the effect. Two types of functions were used to simulate these local effects on the parcels: exponential and negative exponential.

The exponential function has the following form:

$$LEFacg,h = \exp(\theta g \cdot \{ DValueg / RefDValueg \})$$

The negative exponential function has the following form:

$$LEFacg,h = \exp(\theta g \cdot [1 - \{ DValueg / RefDValueg \}])$$

The exponential function is used when the proximity of the local effect increases the rent, while the negative exponential function is useful to represent local effects that decreases the rents as the distance to the effect get smaller.

The form of the function (F (x)) and the expected (Exp.) sign is presented for each space type and for each local effect for the ARC PECAS model in Table 15. Note that there is a maximum distance established for each local effect, meaning that after this point there will be no further effect on the parcel rents. The maximum distances for each local effect are shown in Table 16.

A rent modifier factor needed to be estimated for each local effect and for each space type. This was performed for the ARC PECAS model performing multiple regressions with data on parcel rents and a linear transform of the rent modifier equation. The rent data only included parcels from Fulton County. But, the modifiers were applied to the entire parcel database for the Atlanta Region.

A linear transform of the rent modifier equation is considered in the multiple linear regression estimation, as follows:

$$\ln(Rent_h) = \ln(Price_{h,z}) + \Sigma_{g \varepsilon G} (a_g \times x_{g,h})$$

where:

$x_{g,h}$ = independent variable associated with local-level effect g (g factor)
a_g = coefficient associated with local-level effect g that is estimated in multiple linear regression

A dataset with $x_{g,h}$ is prepared for each local effect g depending on the type of function employed for each effect.

For the exponential function the values are:
xg,h = DValueg / RefDValueg

For the negative exponential function the values are:
xg,h = 1 – { DValueg / RefDValueg }

The regression estimation also takes in account year of the sale, space type, and year built. The resulting coefficients a_g are estimated for each combination of local effects and space types, and are presented in Table 17.

A specific value for $x_{g,h}$ is included for each local-level effect, and the values obtained for the a_g in the estimation are transformed into values for the corresponding θg for the model. The form of the $x_{g,h}$ term to use and the conversion of the associated a_g into θg varies depending on the form of the function. For the functions employed in the ARC PECAS model $\theta_g = a_g$, then no other conversion is required.

The input data with rent observations from the Fulton County and the pre-processing for the multiple regressions to estimate the coefficients are contained in the file called **"Fulton_RentEstimationExample.xls"**. The basic form of the functions and the estimated coefficients are placed in the spread sheet called **"RentEstimationAndLocalEffectsResults-FultonCounty-June2008.xls".** The program to estimate the coefficients was coded in R and it is called **"FultonRentAnalysis.r".**

Table 15. Functional form and expected sign of the local effects by space type for the ARC PECAS model

Local Effects by Space Type	Detached Residential		Higher Density Residential		Industrial		Institutional		Office		Retail	
	F(x) Form	Exp. Sign	F(x) Form	Expected Sign	F(x) Form	Expected Sign	F(x) Form	Expected Sign	F(x) Form	Expected Sign	F(x) Form	Expected Sign
Major Road	Neg Exp	-	Neg Exp	-	Exp	-	Exp	-	Exp	-	Exp	-
Freeway Intersection	Neg Exp	-	Neg Exp	-					Exp	-	Exp	-
Freeway Ramp	Exp	-	Exp	-	Exp	-	Exp	-	Exp	-	Exp	-
MARTA	Neg Exp	+	Neg Exp	+	Neg Exp	+	Neg Exp	+	Neg Exp	+	Neg Exp	+
Schools	Exp	-	Exp	-								
Green Space	Neg Exp	+	Neg Exp	+					Neg Exp	+		

Table 16. Maximum distance associated to each local effect

Local Effect	MaxRoadDist	MaxIntDist	MaxRampDist	MaxMARTADist	MaxSchlDist	MaxSpceDist
Description	Major Road	Freeway Intersection	Freeway Ramp	MARTA (Transit System)	Schools	Green Space
Distance (Feet)	5,280	5,280	7,920	5,280	5,280	7,920

Table 17. Estimated coefficients for each local effect and space type combination

exponential	I (r^2=0.195)				INS (r^2=0.199)			
neg exponential	B	Std. Error	t-ratio	Max (mile)	B	Std. Error	t-ratio	Max (mile)
(Constant)	2.44	0.595	4.098		1.08	0.397	2.719	
Major Road	-0.368	0.126	-2.913	0.5	-0.094	0.171	-0.547	0.5
Freeway					1.504	0.335	4.483	0.25
Freeway Exit	-0.984	0.571	-1.724	0.25	-0.525	0.424	-1.238	5
MARTA	0.993	0.837	1.187	0.25	-0.428	0.214	-2.003	2
School								
GreenSpace					-0.788	0.236	-3.334	1.00
eAge	0	0.003	-0.208		-0.003	0.002	-1.596	

exponential	O (r^2=0.155)				R (r^2=0.153)			
neg exponential	B	Std. Error	t-ratio	Max (mile)	B	Std. Error	t-ratio	Max (mile)
(Constant)	1.403	0.101	13.854		3.525	0.183	19.29	
Major Road	-0.606	0.095	-6.4	0.5	-0.519	0.084	-6.158	0.5
Freeway					-0.052	0.117	-0.442	3.00
Freeway Exit					-0.431	0.176	-2.445	0.25
MARTA	0.713	0.161	4.44	1	0.252	0.128	1.968	0.5
School								
GreenSpace	0.581	0.113	5.14	1.00				
eAge					-0.005	0.001	-4.593	

exponential	RD (r^2=0.194)				RH (r^2=0.298)			
neg exponential	B	Std. Error	t-ratio	Max (mile)	B	Std. Error	t-ratio	Max (mile)
(Constant)	2.775	0.044	63.602		3.009	0.047	63.522	
Major Road	-0.083	0.006	-14.56	0.5	0.056	0.01	5.58	0.5
Freeway	-0.008	0.008	-1.037	3.00	-0.026	0.015	-1.736	3.00
Freeway Exit	-0.282	0.041	-6.95	0.25	-0.027	0.033	-0.81	0.25
MARTA	-0.116	0.011	-10.44	1	0.056	0.013	4.269	1
School	0.045	0.013	3.454	0.25	0.063	0.025	2.532	0.25
GreenSpace	0.04	0.007	5.835	0.25	0.033	0.01	3.428	0.25
eAge	-0.001	0	-12.31		-0.005	0	-42.01	

As examples of the rent estimation functions using the estimated coefficients for some specific local effects on different space types are shown in Figure 9 (for Industrial parcels), Figure 10 (for Institutional parcels), Figure 11 (for high density residential) and in Figure 12 (for residential detached).

Figure 9. Effect of distance to major roads to industrial parcels

Function:

Exponential Coefficient ag: -0.368

	x = Dist	Exp(a(x))
No Dist	0	1.0000
	0.1	0.9639
	0.2	0.9290
	0.3	0.8955
	0.4	0.8631
	0.5	0.8319
	0.6	0.8019
	0.7	0.7729
	0.8	0.7450
	0.9	0.7181
Max Dist	1	0.6921
	1.1	0.6921
	1.2	0.6921
	1.3	0.6921
	1.4	0.6921
	1.5	0.6921

Industrial - Effect: Distance to Major Road

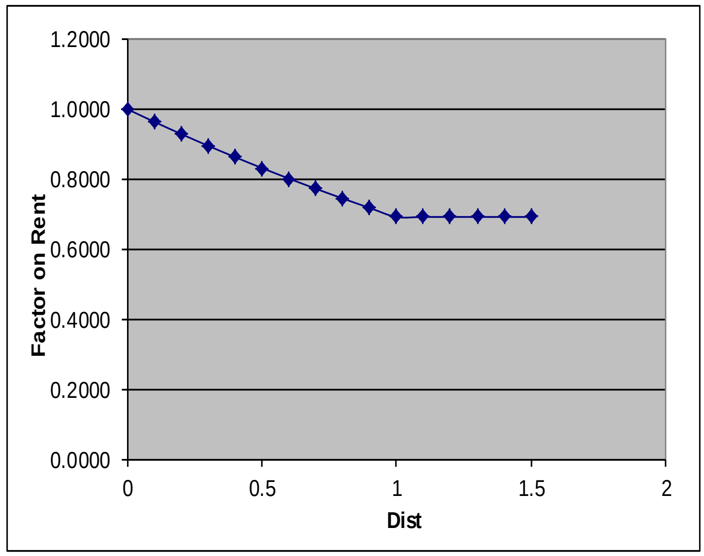


Figure 10. Effect of distance to freeway exit to parcels with institutional space

Function:

Institutional - Effect: Distance to Freeway Exit

Exponential Coefficient ag: -0.525

	x = Dist	Exp(a(x))
No Dist	0	1.0000
	0.1	0.9489
	0.2	0.9003
	0.3	0.8543
	0.4	0.8106
	0.5	0.7691
	0.6	0.7298
	0.7	0.6925
	0.8	0.6570
	0.9	0.6234
Max Dist	1	0.5916
	1.1	0.5916
	1.2	0.5916
	1.3	0.5916
	1.4	0.5916
	1.5	0.5916

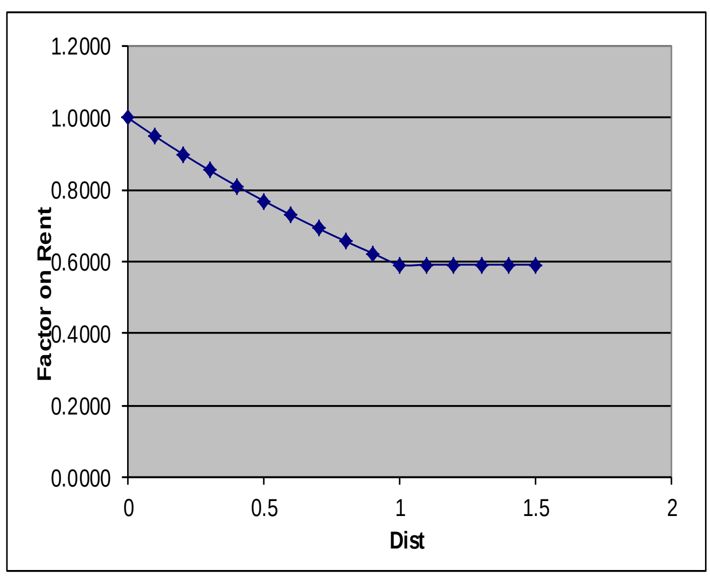


Figure 11. Effect of distance to freeway exit to high density residential parcels

Function:

Negative Exponential   Coefficient ag: -0.026

	x = Dist	Exp(a(1-x))
No Dist	0	0.9743
	0.1	0.9769
	0.2	0.9794
	0.3	0.9820
	0.4	0.9845
	0.5	0.9871
	0.6	0.9897
	0.7	0.9922
	0.8	0.9948
	0.9	0.9974
Max Dist	1	1.0000
	1.1	1.0000
	1.2	1.0000
	1.3	1.0000
	1.4	1.0000
	1.5	1.0000

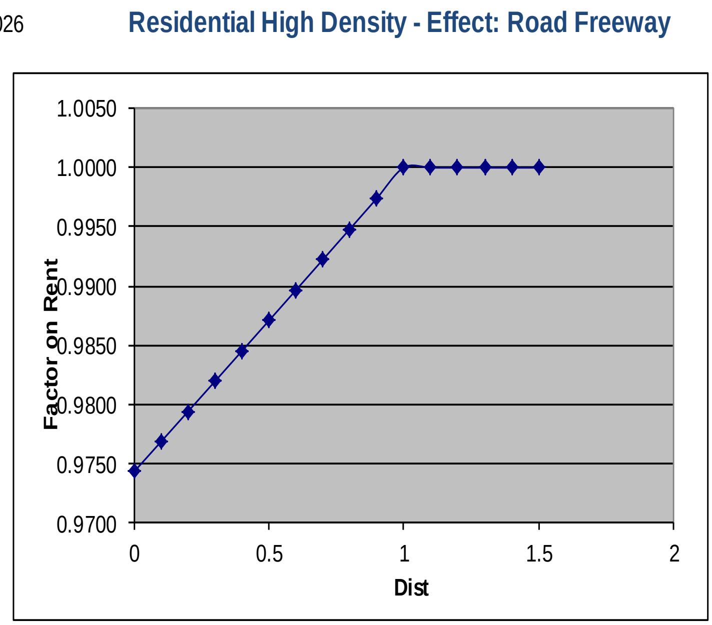
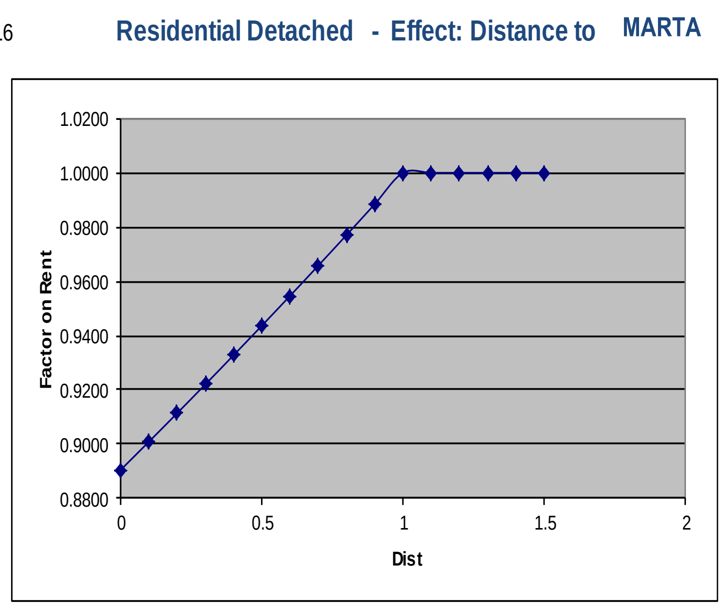


Figure 12. Effect of distance to MARTA to residential detached parcels

Function:

Negative Exponential   Coefficient ag: -0.116

	x = Dist	Exp(a(1-x))
No Dist	0	0.8905
	0.1	0.9009
	0.2	0.9114
	0.3	0.9220
	0.4	0.9328
	0.5	0.9436
	0.6	0.9547
	0.7	0.9658
	0.8	0.9771
	0.9	0.9885
Max Dist	1	1.0000
	1.1	1.0000
	1.2	1.0000
	1.3	1.0000
	1.4	1.0000
	1.5	1.0000



Development Fees & Construction Costs

The PECAS functionality to have development fees that vary across jurisdictions is not used in the ARC PECAS model. Development fees are set to zero, and instead incorporated into the construction costs.

Construction costs for space types were estimated by using a commercial construction cost estimation package called RSMeans (http://www.rsmeans.com) . The software was used interactively to generate Atlanta-specific construction costs for representative developments (typical size and quality.) Construction costs do not vary across the region in the current version of the ARC PECAS model.

Construction Capacity Control (total quantity of construction)

A PECAS feature called *construction control* was implemented in the ARC PECAS model. Construction control allows the SD model's total development to line up with the amount of investment in construction in the financial representation of the industry in the AA module. It works by dividing all of the model's space types into several categories, each limited in a different way by the total of AA's industrial activity. In the ARC model, four categories were used: residential, non-residential, resource, and vacant. The resource category (comprising agricultural and mining space) and the vacant category had no construction control applied, while the residential and non-residential were both tied to the amount of construction production activity (AI04ConProd).

Construction Control is important since otherwise SD views each parcel independently. If some locations in a year become very attractive to developers, those parcels can experience a large quantity of development, which, in reality, would reduce development in other locations in the region, at least temporarily until new capital can be attracted to the region.

Construction control has to be implemented carefully in PECAS, as it can mask consistency problems in the model input data that would be revealed dramatically if SD was allowed to treat each parcel independently. In the ARC PECAS model, while updating to a 2010 base year, and before turning on construction control, it was found that the model was excessively concentrating grade school activity in two zones: north Coweta County (super district 132) and the area around Atlanta International Airport (super district 21). This was traced to an error in the script that was producing the skims files (see the Trip Length Calibration document for more details).

The conversion factor between dollars of construction production and square feet of development was determined by dividing the actual amount of residential and non-residential construction in the base year by the activity total for construction. There was $28 billion of region-wide construction activity in 2005.

7. Overview of procedures to assign space quantities to parcels consistent with the space use rates defined for the Activity Allocation module

The Floorspace synthesizer

Summary

In Section 8 a calibration procedure called "Floorspace Calibration" (page 49) is described, which estimates the amount of floorspace in each zone to best-match a weighted combination of price data and floorspace quantity data. The output of that calibration is an FloorspaceCalc file, which contains the estimated quantity of floorspace-by-type in each zone consistent with the employment and population data, the space use rates that have been established, and the price data, but which is also influenced by the measured amount of space from parcel data. In many counties, the hybrid estimated results in FloorspaceCalc from the calibration is considered a better representation of the built form than the "raw" parcel data. However FloorspaceCalc is established at the TAZ level, and the SD module is a microsimulation module that processes individual parcels.

The Floorspace Synthesizer is a procedure to assign the estimated TAZ level floorspace quantities back onto the parcel data, creating a synthetic parcel database that is similar to the measured parcel database, but that is consistent with the other data sources in the model. With 20 counties in the ARC region, each with it's own tax assessment system and parcel inventory collection, the Floorspace Synthesizer provides consistency between counties and removes the need for county-by-county calibration of the SD module's behavioral parameters.

Details

AA's space consumption functions are a simplification of reality; in particular they cannot represent the observed heterogeneity in the quantity (per employee, or per dollar of production) and category of space use within each industry, even if many industrial categories are used. However, in observed data in the USA, the inconsistency between employment and space extends beyond space-use heterogeneity to data errors, as there are many observations of large non-vacant buildings with no associated employment records, many employment records with no associated buildings, and population totals in census geographies with no housing, and housing records, even non-vacant housing, in geographies with no reported population. Fixing these inconsistencies and identifying errors is possible, but completely fixing them is an infinite task, as new data is collected faster than old data can be cleaned with any reasonable resources.

To address the real world inconsistency, data errors, and heterogeneity, the PECAS AA module is constrained to observed population and employment distributions, then calibrated to best match (in a weighted sum-of-squares of error) observed elastic space use rates, observed quantities of space by zone, and observed floorspace rent (residual zonal rent, adjusted for "local level effects", see section 4) by zone, in the base year.

As a result, AA uses a synthetic representation of base-year built form that is calibrated to best-match observed yet inconsistent datasets that contain measurement errors. This representation is called the Floorspace Inventory.

The SD module simulates development over time on each legal parcel. In PECAS the Floorspace Inventory developed during AA's calibration is often thought to be a more accurate (or at least a more useful and consistent) representation of built form than the detailed parcel data, due to the previously mentioned inconsistencies as well as categorical definitions that differ between local jurisdictions in a region. Thus, a procedure has been developed to assign zonal level Floorspace Inventory to individual parcels in the zone, to best respect observations of parcel conditions.

This approach was applied to most of the counties in the ARC region.

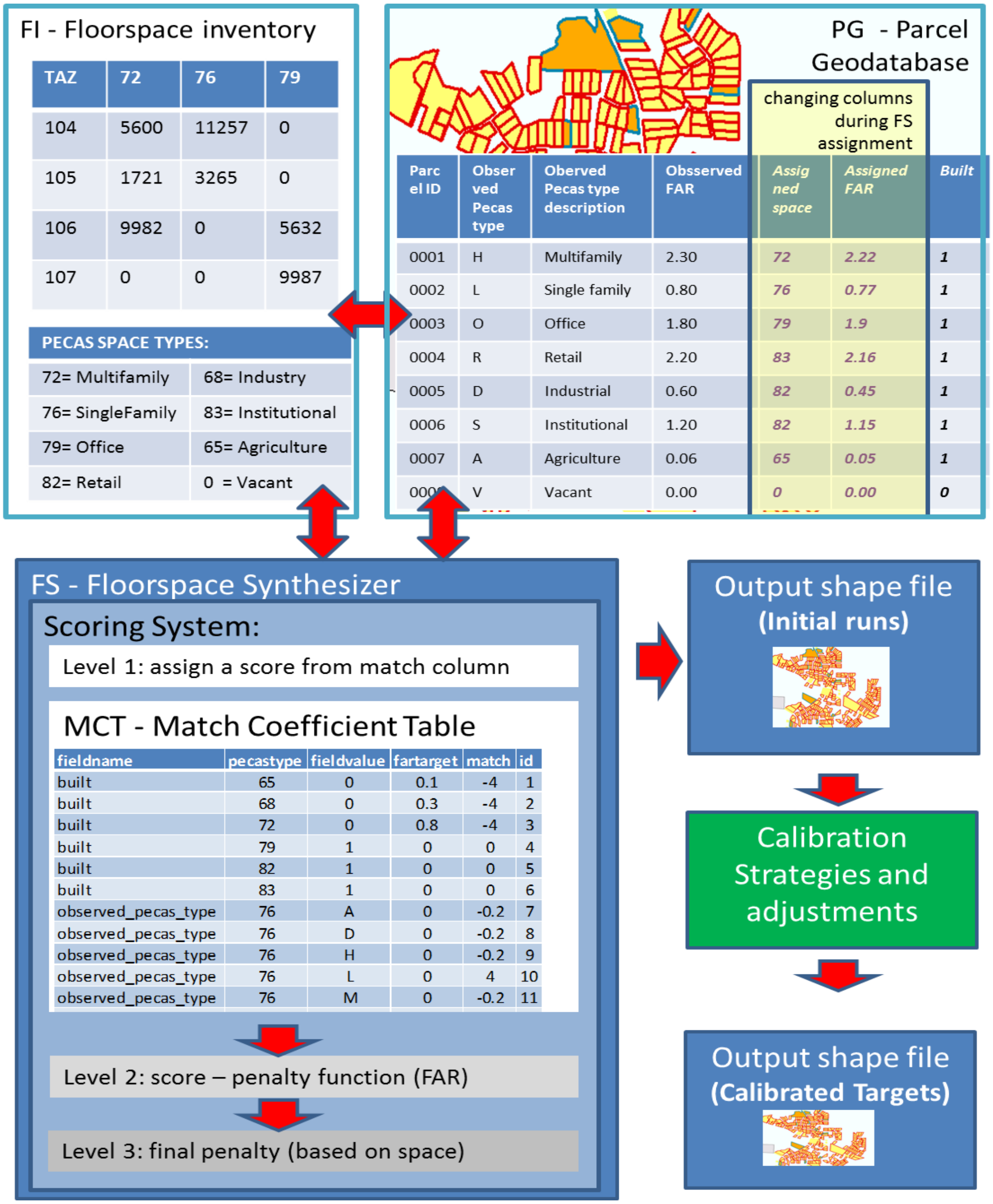
Method applied

The general approach followed is to 1) set up a set of input files, 2) run the synthesizer, examine the results using statistical analysis and map comparisons, and 3) adjust the input files to address concerns. Figure 13 shows the interaction among the floorspace synthesizer components, the steps applied, and categories of space used for the Floorspace Inventory in Clayton (also called the PECAS space types). The four main components needed are:

- The Floorspace Inventory (FI),
- The *Parcel Geodatabase (PG)*, containing attributes of each parcel,
- The Match Coefficient Table (MCT),
- The *Floorspace Synthesizer* (*FS*) program and configuration files

The ***MCT*** is a key component of the ***FS*** mechanism. It defines the ***scoring system*** that is applied during the floorspace assignment to the parcels. Rewards and penalties are established in the ***MCT*** to encourage and discourage types and amounts of floorspace to particular parcels. The ***MCT*** defines the suggestions and hints to be used during the assignment process to indicate whether the parcel is an appropriate location for a type and quantity of floorspace. Observations of the type and quantity of space on each parcel are treated as suggestions, subject to measurement error.

Figure 13. Floorspace synthesizer components and scoring system levels



The *fieldname* column of the ***MCT*** specifies the ***PG*** attribute to be used; the *pecastype* column indicates the type of space under consideration from the ***FI.*** If the specified field in the ***PG*** is identical to the value specified in the *fieldvalue* column, the score indicating the suitability of the space for the parcel is given (or modified) by the value in the *match* column. Similarly, the target for the intensity of development (the floorspace per area of land, called "Floor Area Ratio", or *FAR*) is calculated using this formula:

$$FARTarget_{p,s} = FARTarget_{obs,p} + \sum_{i \epsilon I_s} \delta_{p,i} \cdot FARTarget_i$$

Where,

$FARTarget_{p,s}$	=the FAR target for parcel p and space type s, calculated internally in the FS
$FARTarget_{obs,p}$	=the measured FAR for parcel p in the PG
$\delta_{p,i}$	= 1 if the parcel p matches the criteria specified in row i of the MCT, 0 otherwise
$FARTarget_i$	= the entry in the *FarTarget* column of the MCT for row i
I_s =	the set of rows in the MCT referring to space type s

This method was developed for the ARC PECAS model using data from Clayton County. This procedure was later applied to De Kalb County.

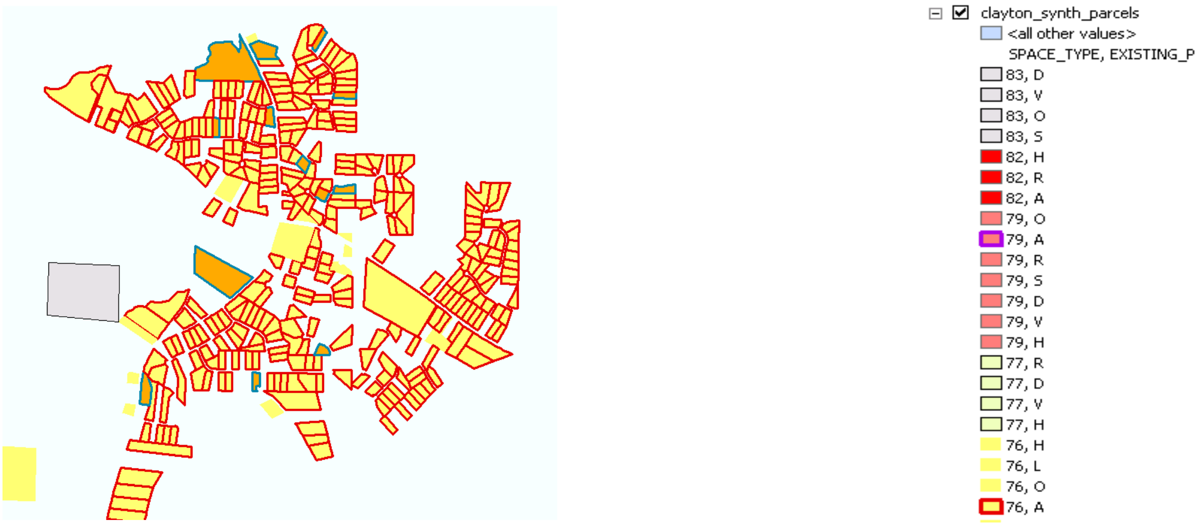
In Clayton County the $FARTarget_i$ entries were only used in the MCT to specify the default intensity targets for the parcels that were observed as vacant (*built*=0). Negative values in the *match* column were used to discourage the assignment to specific parcels, to avoid assigning space to road right-of-ways, known derelict space, and (with a lower penalty) to these parcels where no built space was observed.

The FS calculates an initial score (Level 1 score) for each parcel for each of the PECAS space types in Clayton County. These lists are then sorted, to identify the most suitable parcel for assignment of each type of space. The FS then assigns a small amount of space to the top parcel in each list, iterating through the types of space. With this scoring the best parcels would remain on the top of one list even after they are "full", as indicated by the calculated $FARTarget_{p,s}$. The mechanism to avoid over assigning to individual parcels is Level 2 of the scoring system: a penalty function designed by the analyst that reduces the parcels' scores as their *Assigned FAR* (assigned quantity of space divided by land area) approaches or exceeds the $FARTarget_{p,s}$. Full parcels drop down the (re)sorted lists to less favorable positions and space then gets assigned to other parcels (Figure 14).

A swapping procedure is necessary to ensure that the full inventory can always be assigned without ever assigning multiple space types to any one parcel. A Level 3 modification to the scoring controls this swapping. For details of how to set up these penalties and how they work see a complete report called **Generating PECAS Base Year Built Form for Clayton County in Atlanta.docx** (Fuenmayor and Abraham 2013).

The output GIS layer from the FS includes the original parcel attributes as well as the assigned space type and amount. The primary analysis then involves comparing the observed space type and $FARTarget_{obs,p}$ with the assigned space type and Assigned FAR, using tables, charts, one map for each assigned space type, and one map for intensity of FAR.

Figure 14. Example of parcels with agriculture assigned as single family



The initial run results in Clayton showed that the floorspace was not spread out, leaving many parcels vacant. Adjustments were required in the MCT, in the penalty function (level 2 of the scoring system) and in the PG.

An important part of the calibration process is to adjust the numbers in the *match* column of the MCT in order to reward or to penalize each possible combination of observed space type and PECAS space type. The new penalty function allowed spreading the space on more parcels, while still leaving some parcels vacant.

The lack of spatial detail on dwelling types in the census data (not reported at the block level) was found to lead to some TAZ level inconsistencies that were addressed by identifying multifamily locations in aerial imagery.

After these changes were incorporated, and the synthesizer re-run, some further concerns were identified. For example, many parcels categorized as agriculture were assigned single family dwellings. Parcels observed as having Agriculture were investigated using satellite images, and acreage-style housing was found on many of them. This error in observed space type was forcing the FS to use other land when assigning the known amount of residential housing. (The synthesizer was correct in assigning residential space to parcels that had been observed to have agriculture land; but it had no information to identify which of the "observed agricultural" parcels it should use) (Figure 14).

The MTC was adjusted to give a small boost to the score of parcels under 140,000 square feet for single family space, tagging them as more suitable for residential use than the larger parcels. Further hints were considered (but not implemented) based on proximity to other small parcels, location next to road right-of-ways, orientation vis-à-vis road right-of-ways, etc, all of which might suggest a higher probability of residential (instead of purely agricultural) use.

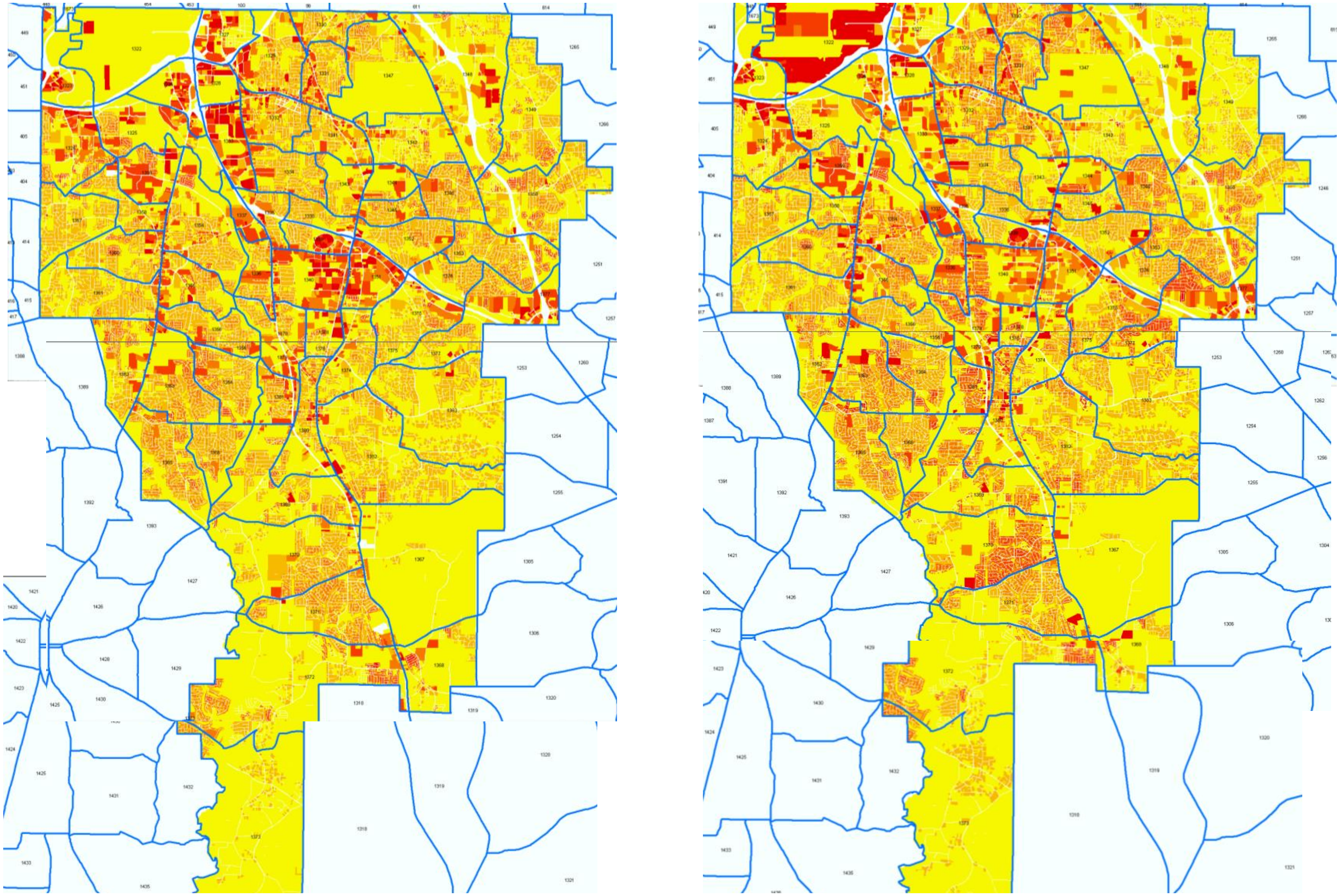
There were other concerns regarding multifamily and office space. Again, reviewing the parcel data in a GIS, together with satellite imagery, and using a GIS analyst's expert judgment, parcels that were more appropriate destinations for the quantity and type of space were identified, and scoring was adjusted. This led to a greater match to observed parcel attributes in cases where the observed attributes seemed correct even after this further investigation.

Major results and improvements

In the final synthesis, the space type assignment in the parcels improved in several ways. The parcels with multifamily achieved a match to observed space type in 94% of the parcels, single family and office had a match of 97%, and vacant parcels showed a 99% match. Lower mismatches in retail (49%), and industrial (23%) space were deemed appropriate, since PECAS space types are categorized by interior function (office, retail and industrial function) while parcel observed space type is largely decided based on exterior form and zoning (a very large number of parcels are observed as industrial).

(Note that 100% matching between assigned floorspace and measured floorspace would suggest that the Floorspace Synthesizer was largely unnecessary, that the amount of space estimated from the FloorspaceCalc.csv file was consistent with the measured space on the parcel database, and that the kinds of problems that typically make the Floorspace Synthesizer necessary are not prevalent in the county.)

A side by side comparison of the observed FAR and the assigned FAR is shown in Figure 15. This visual comparison was reviewed by experts to judge the process, and was considered reasonable. The synthesized version shows more intensity in the airport area and less intensity in other zones and corridors, due to the very large number of employees (and implied very low space per employee) measured at the airport. Special consideration of airport employees may be an appropriate future enhancement to the ARC PECAS model, in the meantime the airport area is consistently represented in the AA and SD modules through increased airport buildings.



Note: Colors represents FAR intensity, from yellow (less intensity) to red (more intensity)

Figure 15. Observed FAR versus simulated FAR

Floorspace Delta tracking SD Parcel Data differences from AA

If the raw parcel database is adequate for SD simulation, the Floorspace Synthesizer (described above) does not have to be used. In this case the AA module and the SD module will have a slightly different view of the amount of floorspace in each zone. The AA module will use the FloorspaceCalc.csv file estimated through the best fit calibration to parcel TAZ-level inventory data and zonal rent data, constrained to population and employement totals, making use of AA's floorspace consumption (floorspace demand) functions. The SD module will simulate future development on each parcel in response to the AA generated prices (based on AAs matching of supply to demand), but will do so while considering the measured state of each parcel.

The PECAS scripting software will notice the presence of a FloorspaceCalc.csv file in any year where AA's floorspace calibration has been performed. It will take the opportunity in that year to create a FloorspaceDelta file, which is a file keyed by TAZ and space type, recording the differences between the FloorspaceO.csv file produced by SD and the calibrated FloorspaceCalc.csv file. The FloorspaceCalc file will be fed to AA as FloorspaceI.csv, so AA uses its appropriately calibrated floorspace.

In future years, the FloorspaceDelta so created will be read by the script. The FloorspaceO file created by SD (representing SD's adjusted view of the development in each TAZ) will be modified by the FloorspaceDelta calculated in the calibration year, and fed to AA as the FloorspaceI.csv file. In this way, the absolute *changes* in TAZ-by-TAZ inventory simulated as development events in SD will be communicated as changes in the FloorspaceI.csv file read by AA (unless changes would lead to a negative amount of space due to a decline in SD inventory greater than the original AA inventory of a type in a zone.)

Note that if the Floorspace Synthesizer *has* been used recently on a county, the entries in FloorspaceDelta for the TAZs in that county will all be 0.0, as the two modules have consistent views of that county. Thus, the FloorspaceDelta functionality is compatible with the Floorspace Synthesizer and does not need to be disable for counties processed by the Floorspace Synthesizer.

The FloorspaceDelta file contains important measures of the consistency between the different data sources used to identify land use and spatial activity for PECAS. However the Floorspace Calibration procedure described below provides a more comprehensive output file describing the same inconsistencies.

8. Overview of procedure followed to calibrate AA to the target data

Generally speaking, the word "calibration" refers to the processes applied to establish parameter values and constants that better allow the model to reproduce the modelled conditions, without compromising any important piece of the model. This requires having target data to represent the conditions to be modelled.

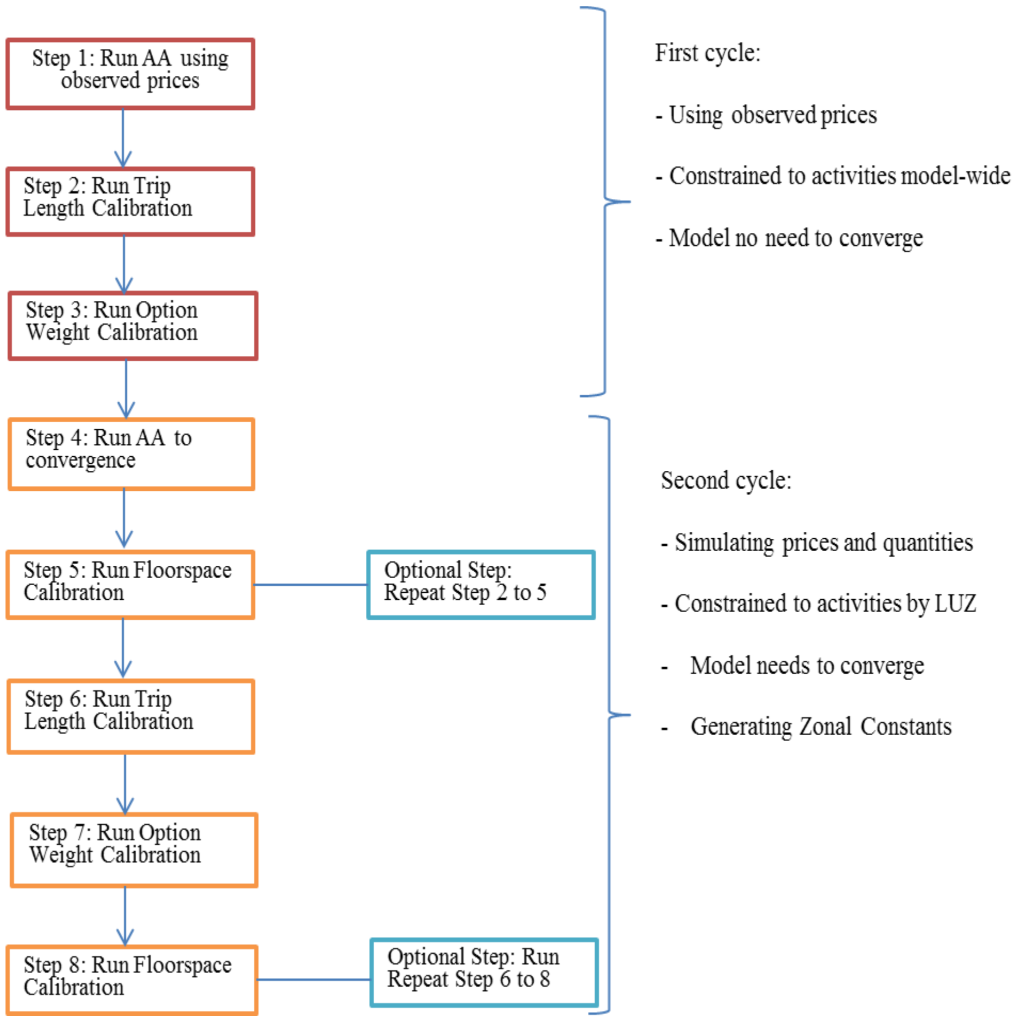
Under the PECAS framework three types of calibration are routinely and systematically performed: trip length calibration, option weight calibration and Floorspace calibration. The parameters and targets associated with each type are presented in Table 18.

Table 18. Parameters and targets associated to each calibration type

Calibration type	Parameter	Targets
Trip Length	Parameter to control the sensitive of buyers and sellers of a commodity c, to differences in desirability between exchange locations	Average trip length for the commodities
Option Weight	This parameter is a term inside of the constant of the utility function for the technology choice	Quantities of households by category with labour occupations and average use of residential space Quantities of non-residential space used by industry Quantities of imports and exports for import providers and export consumers
Floorspace	There is no parameter associated with this calibration. The model adjusts space quantities in order to match rents or adjust rents in order to match space targets.	Quantities of space by type and by zone Estimated rents by zone

Based on experience with several PECAS models, an AA calibration sequence was developed for the PECAS Sacramento model. A series of training videos explain these sequences (J. E. Abraham 2014), and a simplified flowchart is shown in Figure 16. To follow this sequence allows certain components to be calibrated to a certain degree, where the targets are close to being met, and then improvements can be made in all of the components, revisiting some steps with more restrictive conditions. This suggested sequence was used as a general guide for the ARC PECAS model.

Figure 16. Calibration Sequence Recommended for the PECAS AA Module



The following three subsections explains briefly the three types of calibration performed for the AA module and also shows target data and examples of the results from the ARC PECAS models.

The Trip Length calibration

The aim of trip length calibration is to make the average length of the commodity flows generated by AA match the actual average trip length for that type of commodity. This is done by adjusting the buying and selling dispersion parameters for each commodity until the average trip lengths equal the targets.

These dispersion parameters appear in the formulas for allocating the buying and selling of commodities to the different exchange zones. Since the dispersion parameter is multiplied by the utility values for exchange size, commodity price, and generalized cost of transportation, it affects the way that the exchange zone allocation responds to those signals. If the dispersion parameter is large, the allocation becomes more sensitive to transport cost, meaning that AA will exchange more of the commodity in nearby zones, reducing the average length of trips by that commodity.

The trip length calibration script (TLC.py) matches the targets by increasing the dispersion parameters for commodities whose trips are too long and decreasing them for commodities whose trips are too short. The TLC algorithm does this by running the AA model several times by adjusting the dispersion parameters until it matches targets. In ARC PECAS model, travel length is affected by the transport disutility of the 22 transportable commodities defined for this region. The commodities were divided in to 8 groups. The assigned groups of commodities and their travel time targets are shown in Table 19.

The targets for the average trip length are reported in minutes and the data was obtained from the trip length frequency distribution by trip category and income, produced by the calibrated trip model for the Atlanta Region, which is the previous version of the current Activity Based Model.

Before the TLC script is run, two files are needed to indicate the setup of the calibration. TLCGroupsI.csv has the groups of the commodities, while TLCTargetsI.csv which has the initial dispersion parameters and target trip length values. Moreover, the number of maximum iterations and allowable error for the trip length should be specified in the script. The skim matrix used to calibrate the parameters of each commodity and intervals of the trip lengths used to calculate the average trip lengths are specified in HistogramsI.csv.

When TLC.py runs, it creates the TLCCalib.csv which has the results of the average trip lengths, estimated parameter and model error for every iteration. When the trip length error of all commodities is less than five percent, the run will be stopped and the trip length for each commodity for given trip length intervals are written in the file called Histograms.csv.

Table 19. Commodities, Groups and Target distance values used for the trip length calibration

Commodity	Group	Target (Average Trip Time in minutes)
CG01AgMinDirection	Work-Other	18
CG02AgMinOutput	CV Heavy	45
CG03ConDirection	Work-Other	18
CG04ConOutput	CV Heavy	45
CG05MfgDirection	Work-Other	18
CG06MfgOutput	CV Med/Light	35
CS07TCUDirection	Work-Other	18
CS08TCUOutput	CV Med/Light	35
CS09WsOutput	CV Med/Light	35
CS10RetailOutput	HB Shop	16.4
CS11FIREOutput	Work-Other	18
CS13OthServOutput	HB Other	17.5

Commodity	Group	Target (Average Trip Time in minutes)
CS14HealthOutput	HB Other	17.5
CS15GSEdOutput	HB School	15.48
CS16HiEdOutput	HB Univ	28.95
CS17GovOutput	Work-Other	18
CL23WhiteCollar	HB Work	34
CL24Services	HB Work	34
CL25Health	HB Work	34
CL26Retail	HB Work	34
CL27BlueCollar	HB Work	34
CL28Military	HB Work	34

The Option Weight calibration

Option size calibration is the second step in calibrating the AA module of PECAS, following trip length calibration.

PECAS uses technology options (TechnologyOptionsI) to represent the different combinations of commodities that an activity can produce and consume. The aim of option weight calibration is to find weights such that the total amounts of production and consumption for certain commodities match targets. The option weight is part of the constant, one of the terms in the technology utility for each option. Changes in the constant affect the probability of choosing that technology option and ultimately impact the total quantity chosen (produced or consumed) from that option.

The option weight term for a technology option reflects the "inherent desirability" of that option relative to the others. With all else equal, the odds of choosing a technology option is proportional to that option weight term.

There can be number of options in a same activity. In PECAS, tags in the TechnologyOptionsI file are used to indicate how each technology option differs from the others under the same activity. With respect to a single commodity, option size calibration relies on three tags for each option describing how it relates to levels of production or consumption of commodities: "more", "less", or "zero".

Option weight calibration adjusts the weights for each technology option to match production and consumption targets. Each target is the total amount of a given commodity that should be made or used by a given activity across the entire model region.

Option weight calibration uses an iterative approach, each iteration adjusting the weights on "more" and "less" options to move the aggregated production and consumption amounts towards the targets. The adjustment factor for "less" options is smaller than that for "more" option: both "less" and "more" options contribute to the overall production or consumption of a commodity and so they both need to be adjusted. (This means that even if one or more of the options is missing e.g. there is a “more” and a “less” option but no “zero” option, the balance between the options will still shift in the right direction.).

Option weight calibration is usually performed in two stages. In the first stage, calibration is done unconstrained with no equilibrium search in AA, in order to reach a region-wide balance of production and consumption at prices that are exogenously set to be close to observed prices. In the second stage, calibration is done constrained with AA run to convergence, which achieves a balance with equilibrium zone-by-zone prices. During this second stage, AA was set to not change the prices of the floorspace commodities, as those were treated in the floorspace calibration step instead.

A program to automatically adjust the option weight parameters until it matches the targets reported in the file called OptionSizeCalib.I.csv. Before the program is run, the make and use targets for the model area needs to be specified. Other settings (e.g. the number of maximum iteration) can be specified in the file option_size_calib.py.

At each stage of the option weight calibration, the total absolute error in the amounts produced and consumed by each activity were calculated before and after the calibration run and they are written in the file called OptSizeCheck.csv file. Table 20 summarizes these results, showing the progression towards the calibration targets. The household activities are highlighted, since they needed the greatest adjustment (including manual adjustment) to match their targets; in fact, most of the other activities already matched their targets due to previous calibrations.

Table 20. Summary of progress across all stages of option weight calibration

		Stage 1 - % Error				Stage 2 - % Error	
Activity	Target ($)	Initial	First Run	Manual	Second	Initial	Final
AH29HHlt20_12	2.51E+09	60%	45%	100%	7%	45%	1%
AH30HHlt20_3p	1.09E+09	52%	33%	98%	5%	8%	6%
AH31HH2050_12	1.55E+10	49%	22%	30%	4%	15%	2%
AH32HH2050_3p	1.19E+10	42%	6%	5%	1%	3%	0%
AH33HH50100_12	2.53E+10	2%	0%	0%	0%	14%	0%
AH34HH50100_3p	3.44E+10	2%	0%	0%	0%	13%	1%
AH35HHge100_12	2.14E+10	1%	0%	0%	0%	7%	1%
AH36HHge100_3p	3.18E+10	0%	0%	0%	0%	6%	1%
AI01AgMinMan	1.62E+08	0%	0%	0%	0%	0%	0%
AI02AgMinProd	1.68E+08	0%	0%	0%	0%	0%	0%
AI03ConMan	3.76E+09	0%	0%	0%	0%	0%	0%
AI04ConProd	6.16E+09	0%	0%	0%	0%	0%	0%
AI05MfgMan	6.81E+09	0%	0%	0%	0%	0%	0%
AI06MfgProd	6.04E+09	0%	0%	0%	0%	0%	0%
AI07TCUMan	6.11E+09	0%	0%	0%	0%	0%	0%
AI08TCUProd	6.69E+09	0%	0%	0%	0%	0%	0%
AI09Whole	1.30E+10	0%	0%	0%	0%	0%	0%
AI10Retail	1.22E+10	0%	0%	0%	0%	1%	0%
AI11FIRE	1.50E+10	0%	0%	0%	0%	0%	0%
AI12PTSci	1.65E+10	0%	0%	0%	0%	0%	0%
AI13ManServ	4.59E+09	0%	0%	0%	0%	0%	0%
AI14PBSOff	1.04E+10	0%	0%	0%	0%	0%	0%
AI15PBSRet	4.13E+09	0%	0%	0%	0%	0%	0%
AI16PSInd	3.04E+08	0%	0%	0%	0%	0%	0%
AI17Religion	9.93E+07	0%	0%	0%	0%	1%	0%
AI18BSOnsite	4.46E+09	0%	0%	0%	0%	0%	0%
AI19PSOnsite	3.46E+08	0%	0%	0%	0%	0%	0%
AI20FedGov	4.75E+09	0%	0%	0%	0%	0%	0%
AI21StLocGov	3.69E+09	0%	0%	0%	0%	0%	0%
AI22Military	1.09E+09	0%	0%	0%	0%	0%	0%
AI23GSEdu	6.99E+09	0%	0%	0%	0%	0%	0%
AI24HiEdu	3.00E+09	0%	0%	0%	0%	0%	0%
AI25Health	9.19E+09	0%	0%	0%	0%	0%	0%
Exporters	1.61E+11	0%	0%	0%	0%	1%	0%
Importers	1.34E+11	0%	0%	0%	0%	4%	0%

More detail description of the Option Size or Option Weight calibration is available in the technical document called "**Option Size Calibration ATLANTA.docx**" file.

Floorspace calibration

Approach

Floorspace calibration is the process of adjusting the quantity of space to achieve prices that better match observations; normally, adding space of a given type will decrease the price of that space, and vice versa.

The reasons for this are discussed previously in section 7. Essentially there are two reasons: 1) AA uses simplified floorspace demand functions that do not account for real-world heterogeneity in space use rates, and 2) in the United States context measured floorspace quantities have always shown inconsistencies with employment and population data (usually due to mistakes in data collection in the Census data or Parcel data, or inconsistent time periods, e.g. a new subdivision where new residents were not identified until some future-year data collection).

Adjusting the floorspace quantities to exactly match the prices used to be the recommended modelling approach, however some time ago (in a different PECAS project) it was identified that this would severely distort the distribution of different types of space in cases of substitutable space types. Given that both quantity and price data contain both errors and interpretation mismatches, and that the model itself is imperfect, floorspace calibration attempts to balance between matching the observed quantity and price, while respecting spatial distributions of activity and floorspace consumption functions. To this end, a "tolerance" value is assigned to both the quantity target and the price target for each LUZ/space type combination, representing the uncertainty of that target. If, for example, the tolerance for the price target is high while that for the quantity target is low, then after calibration, the quantity will match its target more closely than the price matches its target.

The details about the mathematical approach of the Floorspace Calibration are contained in a technical documentation called "Theoretical and Mathematical approach of the Floorspace Calibration" (Hill 2012).

Target data

Space quantity targets were derived from the model's 2005 floorspace inventory; see section 6 of this document for details of how the floorspace inventory was constructed.

The tolerance values were simply half the floorspace quantity (e.g. 50,000 square feet in a zone containing 100,000 square feet), to a minimum of 10,000 square feet. The minimum tolerance assigned only when the half of the space in a zone is less than the minimum quantity.

Price targets were derived from the space rent estimation described in section 4 of this report.

Results

Since floorspace calibration is a balance between matching the quantity and price, its success is measured using the *total error*. For a given type of space in a given zone, the error for that combination is given by the following equation:

$$Error = \frac{(ModelledPrice - TargetPrice)^2}{PriceTolerance^2} + \frac{(ModelledSpace - TargetSpace)^2}{SpaceTolerance^2}$$

These error values can be aggregated to produce various total error measurements. The total error across all zones for each space type, both before and after calibration, is shown in Table 21.

Table 21: Total error by space type before and after floorspace calibration

Space type	Total error before	Total error after
Agriculture & Mining	0.0	0.0
Industrial	0.0	0.0
Retail	3.6	0.8
Office	1.4	0.2
Institutional	0.1	0.1
Detached Residential	40.3	17.5
High-Density Residential	85.4	17.0

Figures 17 and 18 show more detail about how the individual zones changed because of floorspace calibration. In each graph, the target price (in dollars per square foot per year) appears on the horizontal axis, while the modelled price appears on the vertical axis. Each circle represents the price of a certain space type in one zone. The black diagonal line represents a perfect match; if the model was perfect, or if price match was given an extremely low tolerance as compared to the floorspace quantity targets, all circles would lie exactly on this line.

The non-residential space types appear in Figure 17. Even before the calibration, all of the circles were very close to the ideal line, indicating that the prices were already reasonable. After calibration, they were even closer.

Figure 17: Fit to price targets – non-residential space

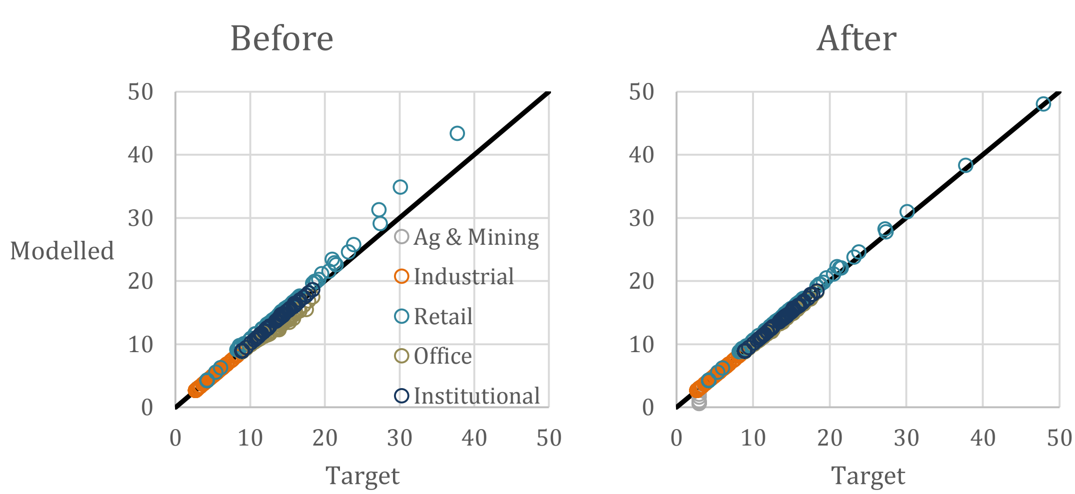
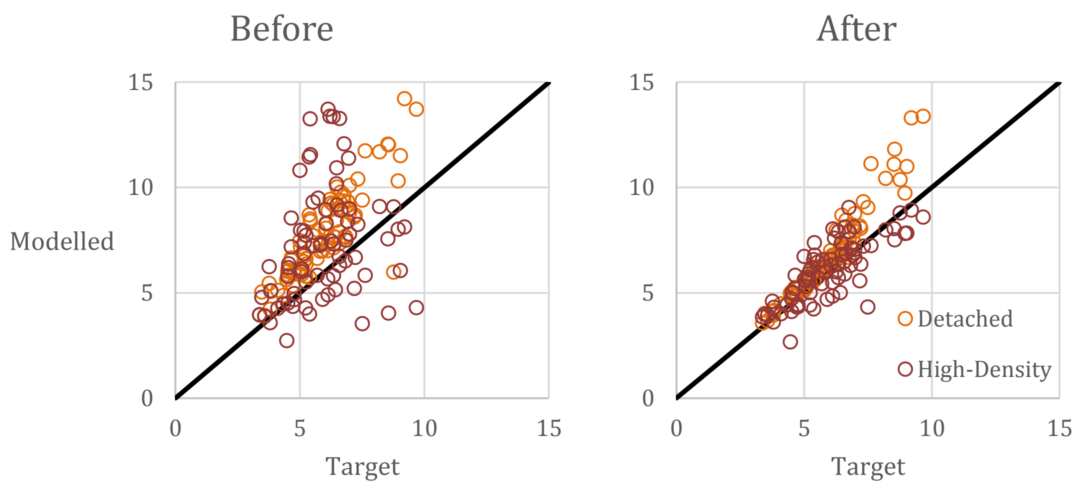


Figure 18 shows the price of residential space. Both detached and high-density residential space showed significant improvements in matching the observed prices. Before calibration, the price of detached residential space was consistently high, while high-density residential had a weak correlation between the observed and modelled price in a zone. Both of these shortfalls were addressed by floorspace calibration, though the final fit was not perfect because of the competing need to avoid large changes in space quantity.

Figure 18: Fit to price targets – residential space



9. Overview of procedure followed to calibrate the Spatial Development module to historical data of development

Approach

The space development (SD) module of the PECAS framework has the objective of simulating development at the parcel level taking into account developer behavior.

The core of the Space Development (SD) model is based on rational profit maximization – developers are more likely to build where construction costs are low and rents are high. However, some parameters are left unspecified and must be estimated based on observations of real activity. Most of these parameters are market share constants: they add a fixed value to a particular alternative under certain conditions to align the total share of developers choosing that alternative with reality. These constants are designed to represent unknown influences on decisions, especially ones that are not of interest for policy analysis. The objective of SD calibration is to find values for these constants and other parameters that best represent the actual behavior of the land development system.

SD calibration is a "Bayesian" method – it combines prior information about the likely parameter values with new information. This is important because development displays a high degree of variability due to its discrete nature, as entire office towers or neighborhoods are built or not built based on changing economic conditions. Also, some types of development are rare, so the data quality for those types is lacking. The prior information helps temper these variations, preventing the parameters from being influenced excessively by anomalous data and providing a fallback in place of absent data.

SD calibration is an iterative approach that converges towards the most likely parameter values, which are those that reproduce the target data as well as possible without drifting too far from their prior values. As with floorspace calibration, "tolerance" values are specified that determine the balance between matching different targets and priors. Since SD is non-deterministic (and so outcomes vary from one run to the next), each iteration calculates the expected value of each outcome so that there is consistency between iterations.

A detailed explanation regarding the Bayesian Calibration Software for the SD Module including the setup and the output description is contained in a technical document called SD Calibration (Hill 2014).

Target data and priors

The priors were generally "low-information"; they had large tolerances so that they would not unduly influence the calibration, but would provide a reasonable fallback where little information about that parameter could be gained from the target data.

The target data for SD calibration came from F.W. Dodge construction reports. Records of new construction were aggregated by county and space type to create targets for both the total quantity and density of development. Addition and renovation targets for residential and non-residential space were also compiled.

Tolerance values were based on judgment of the reliability of the Dodge targets. The development quantity targets (tottarg and grouptarg) were given a fixed tolerance of 100,000 square feet. The addition and renovation targets had tolerances equal to 10% of the target value. For non-residential space, the FAR targets had a fixed tolerance of 0.05; for residential space a larger value was used, with 0.2 for low density and 0.5 for high density.

This step in the calibration has been superseded by the more recent SD recalibration. Details on how the targets and priors were updated, as well as the results of the recent SD calibration, are given in the report called **"The ARC PECAS Model Improvements, current use and ABM integration.docx".**

10. References

Abraham, J. E. 2014. *Calibration Sequence for the Sacramento PECAS Model*. Calgary, AB. http://www.hbaspecto.com/pecas/pecas_videos/.

Abraham, John E., Kevin J. Stefan, and J. D. Hunt. 2012. "Population Synthesis Using Combinatorial Optimization at Multiple Levels." In . http://trid.trb.org/view.aspx?id=1130260.

Fuenmayor, Geraldine, and John E. Abraham. 2013. "Generating PECAS Base Year Built Form for Clayton County in Atlanta." HBA Specto Incorporated.

Hill, Graham. 2012. "Theoretical and Mathematical Approach of the Floorspace Calibration." HBA Specto Incorporated.

———. 2014. "Bayesian Calibration Software for the SD Module." HBA Specto Incorported.

Appendix 1: Mathematical equations to calculate rents for the residential space

To determine location rent by census tract, as well as adjustment factors for housing size, housing age, and housing structure type,

minimize

$$SSE = \sum_{b \in B} \left(pgr_b - ogr_b \right)^2$$

where

SSE = (dis)merit measure to be minimized, sum of the squared errors
b = a block group identifier
ogr_b = observed gross rent for the block group b (variable H64 from Census SF3, in Atlanta, column agg_grnt)
B = set of block groups in the region
pgr_b = predicted gross rent for block group b, given by

$$pgr_b = \sum_{h \in H} \left(RA_b \cdot AA_b \cdot RD_{h,b} \cdot S_h \right) \cdot TR_t$$

h = one of the housing types in the model identifiable from census "units in structure" variable
H = the set of housing types in the model identifiable from census "units in structure" variable (in Atlanta, "1 unit detached" or "other")
RA_b = Number of rooms adjustment factor for block group b, defined below
AA_b = Age adjustment factor in block group b, defined below
$RD_{h,b}$ = Number of rented dwellings of housing type h in block group b (variable H32 from Census SF3), h is defined in census from "units in structure" variable (in Atlanta rent_1_det, or rent_allother)
S_h = Structure type adjustment factor (parameter to be estimated) (in Atlanta rent_1_det, or rent_allother)
TR_t = the tract rent constant for each census tract t, where b is contained within t, parameter to be calculated

With

$$RA_b = \frac{\sum_r RA^r \cdot RD_b^r}{RD_b} \text{ if } RD_b > 0\text{, 1 otherwise}$$

Where
RD_b = Number of rented dwellings in block group b (variable H7 from Census SF3) (it Atlanta, rent_hu)
RD_b^r = Number of rented dwellings in block group b with number of rooms r (variable H26 from Census SF3) (in Atlanta, rent_1r, rent_2r, … rent_9r)
RA^r = Rent adjustment factor for r rooms, given by

$$RA_r = 1 + \phi_1 (r-5) + \phi_2 (r-5)^2 + \phi_3 (r-5)^3$$

ϕ_1, ϕ_2, ϕ_3 = Rent adjustment for rooms parameters (to be estimated)

AA_b = Age adjustment factor in block group b, given by

$$x + (1-x)(1-\alpha)^{(BR - YB_b)}$$

YB_b (in Atlanta, Mdyblt_rnt)	=	Median year built for block group *b* for rented dwellings (variable H37 from Census SF3)
x	=	Portion of housing value not affected by age (parameter to be estimated)
□	=	the age decline parameter (parameter to be estimated)
BR	=	base year for the rent data (e.g. 2000 for Census 2000 data)

By adjusting

□□□□□□□	the adjustment factors for size of dwelling
S_h	the adjustment factors for dwelling type, for $h \neq 1$, i.e. not Single Family Dwelling (S_1 is set to 1.0)
TR_t	the tract rent constant for each census tract *t*, where *b* is contained within *t,* (interpretable as the predicted rent for a 5 bedroom brand new single family dwelling)
□	the age decline parameter
x	the portion of housing value that is independent of age of structure

Subject to

□ >= 0

$TR_t > 0$ for all t

x >= 0

$S_h > 0$

Solving for TR_t directly:

The above formula is linear in TR_t so the TR_t can be solved for directly by the first order conditions,

$$\frac{\partial SSE}{\partial TR_t} = \frac{\partial \sum_{b \in B} \left(pgr_b - ogr_b \right)^2}{\partial TR_t} = 0 \quad \text{at the minimum}$$

so

$$\frac{\partial SSE}{\partial TR_t} = \partial \sum_{b \in B_t} 2\left(pgr_b - ogr_b \right) \frac{\partial pgr_b}{\partial TR_t}$$

where

B_t	=	the set of block groups within census tract *t*

let

$$ABRA_b = \sum_{h \in H} \left(RA_b \cdot AA_b \cdot RD_{h,b} \cdot S_h \right)$$

That is

$ABRA_b$	=	the aggregate block group rent adjustment factor for block group *b*

Then from above

$$pgr_b = ABRA_b \cdot TR_t$$

and

$$\frac{\partial SSE}{\partial TR_t} = \partial \sum_{b \in B_t} \left(2\left(ABRA_b \cdot TR_t - ogr_b \right) \cdot ABRA_b \right) = 0 \quad \text{at the minimum}$$

Solving for TR_t

$$\sum_{b \in B_t} \left(ABRA_b \cdot TR_t - ogr_b \right) ABRA_b = 0$$

$$\sum_{b \in B_t} \left(ABRA_b{}^2 \cdot TR_t \right) = \sum_{b \in B_t} ogr_b \cdot ABRA_b$$

$$TR_t = \frac{\sum_{b \in B_t} ogr_b \cdot ABRA_b}{\sum_{b \in B_r} ABRA_b{}^2}$$

Thus for calculating the *SSE* given parameters [illegible] S_h, [illegible] and x, first calculate $ABRA_b$ for each block group b, then calculate the average tract rents TRt, then calculate the *SSE*.

In this way the numerical non-linear search algorithm only has to determine 6 parameters, regardless of the number of census tracts in the region.

Appendix 2: List of files containing procedures and estimations employed in the development of the ARC PRCAS Model

Subtopic	File name	Description
Developing the Aggregate Economic Flow Table AEFT		
Money Flows between activities and commodities	"PecasIMPLANtemplt_201204. xls".	Data and the procedure to develop the AEFT for the ARC PECAS model
Households providing labour	"PumsProcessing.sql".	The script develop for PUMS processing (labor make)
Categorizing households	"build_samples.sql".	The script for processing households based on income, size and building type
Households using residential space	"pums_hh_sqft.sql	The coefficients and the equations employed for each housing type
Building the entire population of households	"To build targets with Entropy.xlsx"	Targets to match during the population synthesizer process
Labor using non-residential space	"EmploymentAllocationV1B-noSpaceInputs.xlsx"	Calibrating the exponent to split labor based on the space type used
		Calibrating space use rates for non-residential space
Estimating Rents		
Residential Space	"PECAS_BGs_Rents_Base Data.xlsx",	The spread sheet and script for the residential rent estimation
	"RentRegression.py"	The script develop for the parameter estimation for the **residential rent** estimation by census track
	Appendix 1 of this report	The mathematical formulas involved in the rent estimation for the residential space
Non-residential Space	"NonResidentialRentEstimation .xls"	The spread sheet for the rent estimation for the non-residential space
Transport Cost Coefficients		
Transport cost coefficient	"ARC PECAS.TCC.2.xls"	Transport cost coefficient estimation
Parcel data		
Zoning	"zoning.xlsx"	Processing zoning codes
Zoning at regional level	"ZoningForImage.qgs"	Zoning codes at the regional level
Rent modifier based on local effects	"Fulton_RentEstimationExample.xls".	Input data with rent observations for the multiple regressions
	"RentEstimationAndLocalEffectsResults-FultonCounty-June2008.xls".	Basic form of the local effects functions and estimated coefficients
	"FultonRentAnalysis.r"	Program to estimate the coefficients

Subtopic	File name	Description
Floorspace synthesizer and Floorspace Inventory		
Floorspace Inventory	"Generating PECAS Base year built form for Clayton County in Atlanta.docx"	This report documents how to set up the floorspace synthesizer for the ARC PECAS model using Clayton County data as an example.
AA calibration		
Trip Length calibration	"Trip Length Calibration ATLANTA.docx".	The Trip Length calibration for Atlanta
Option Size calibration	"Option Size Calibration ATLANTA.docx"	The Option Size calibration for Atlanta

Appendix 3: Procedure to estimate the regression coefficients and constants to update the transport utility functions for the commodities using data from the ABM model

Introduction

In this section, regression equations were used to predict the old skims of different commodities using the new ABM skims are presented. Moreover, the criteria which were chosen to select the best prediction model and estimated regression model parameters are also documented here. Two separate model types were used for the non-labour commodities (*worker delivery services* and *household obtained services*) and for the labor commodities. They are summarized below.

Model 1

Skims of seven commodities were predicted using the ABM skims. The predicted skim and the ABM skim variables which were analyzed to use in prediction model are shown in Table 1. Two equations which were used as the initiation regression models for two commodities are shown in Eq. 1 and Eq. 2. Forward stepwise regression was used to select the optimum models which have good prediction capabilities. “Leaps” package of R was used for estimate the optimum models. It has a function to perform an exhaustive search for the best subsets of the variables for predicting y in linear regression, using an efficient branch-and-bound algorithm. All the output models were not considered to selection of the optimum variables. The equations which have different coefficient signs were eliminated from the selection. Then the optimum model was selected based on the magnitude of correlation coefficient. The model which has the maximum correlation coefficient was chosen as the prediction model. Table 2 shows the variables selected for optimum model, estimated coefficients and correlation coefficient value for each model.

$y = CG_k + \sum_{i=1}^{n} \beta_i * CG_i$ [Eq. 1]

$y = CS_k + \sum_{i=1}^{n} \alpha_i * CS_i$ [Eq.2]

CS_k, CG_k = Constants

Table 1: Variables used in the regression equations

Commodity (y)	Old skim	ABM skims	Variable in Eq
CG01AgMinDirection CG03ConDirection CG05MfgDirection CS07TCUDirection	LOGSUM4_NHB	op_medInc_autos_lt_workers	CG1
		op_highInc_autos_ge_workers	CG2
		op_veryHighInc_autos_ge_workers	CG3
CS11FIREOutput CS17GovOutput	LOGSUM4_NHB	op_lowInc_0_autos	CS1
		op_lowInc_autos_lt_workers	CS2
		op_lowInc_autos_ge_workers	CS3
		op_medInc_0_autos	CS4
		op_medInc_autos_lt_workers	CS5
		op_medInc_autos_ge_workers	CS6
		op_highInc_0_autos	CS7
		op_highInc_autos_lt_workers	CS8
		op_highInc_autos_ge_workers	CS9
		op_veryHighInc_0_autos	CS10
		op_veryHighInc_autos_lt_workers	CS11
		op_veryHighInc_autos_ge_workers	CS12
CG10RetailOutput CS13OtherServOutput CS14HealthOutput CS15GSEdOutput CS16HiEdOutput	LOGSUM4_HBO	op_lowInc_0_autos	CS1
		op_lowInc_autos_lt_workers	CS2
		op_lowInc_autos_ge_workers	CS3
		op_medInc_0_autos	CS4
		op_medInc_autos_lt_workers	CS5
		op_medInc_autos_ge_workers	CS6
		op_highInc_0_autos	CS7
		op_highInc_autos_lt_workers	CS8
		op_highInc_autos_ge_workers	CS9
		op_veryHighInc_0_autos	CS10
		op_veryHighInc_autos_lt_workers	CS11
		op_veryHighInc_autos_ge_workers	CS12

Table 2: Estimated variables for Eq 1 and 2

Commodity (y)	Old skim	ABM skims	Est. Coefficient	Regression Coefficient (R)
CG01AgMinDirection CG03ConDirection CG05MfgDirection CS07TCUDirection CS11FIREOutput CS17GovOutput	LOGSUM4_NHB	CG_k	0.115	0.947
		CG3	0.687	
CS11FIREOutput CS17GovOutput	LOGSUM4_NHB	CS_k	0.112	0.947
		CS1	0.002	
		CS9	0.681	
CG10RetailOutput CS13OtherServOutput CS14HealthOutput CS15GSEdOutput CS16HiEdOutput	LOGSUM4_HBO	CS_k	0.422	
		CS1	0.00052	
		CS9	0.6333	0.955

Model 3

Second model was used to estimate the interchange coefficients which need to be used in Commodities I file for labor commodities. Table 2 shows the variables which were used to estimate the transport exchange coefficients and the ABM skims which were used as the predicting variables. Eq. 3 shows the equation which was used for the estimation. Linear regression was used to estimate the intercept and coefficient of the dependent variable (ϕ). Estimation results of each labor category is shown in Table 4.

$$\sum_{i=1}^{3} I.C_i * O.S_i = LB_c + \emptyset * (p_1 * LB1 + p_2 * LB2 + p_3 * LB3 + p_3 * LB3 + p_4 * LB4) \text{[Eq.3]}$$

LB_c = Constant
P_i = Proportion of workers with respect to income level using PUMS data
$I.C_i$ = Interchange coefficients
$O.S_i$ = Old skim
LB_i = ABM Skims (Labor)

Table 3: Model 2 variables

Commodity	ABM skims	Variable in the Eq
white collar	op_lowInc_autos_ge_workers	LB1
	op_medInc_autos_ge_workers	LB2
	op_highInc_autos_ge_workers	LB3
	op_veryHighInc_autos_ge_workers	LB4
services	op_lowInc_autos_ge_workers	LB1
	op_medInc_autos_ge_workers	LB2
	op_highInc_autos_ge_workers	LB3
	op_veryHighInc_autos_ge_workers	LB4
health	op_lowInc_autos_ge_workers	LB1
	op_medInc_autos_ge_workers	LB2
	op_highInc_autos_ge_workers	LB3
	op_veryHighInc_autos_ge_workers	LB4
retail and food	op_lowInc_autos_ge_workers	LB1
	op_medInc_autos_ge_workers	LB2
	op_highInc_autos_ge_workers	LB3
	op_veryHighInc_autos_ge_workers	LB4
blue collar	op_lowInc_autos_ge_workers	LB1
	op_medInc_autos_ge_workers	LB2
	op_highInc_autos_ge_workers	LB3
	op_veryHighInc_autos_ge_workers	LB4

Table 4: Estimated variables for model 2

Labor category	Intercept	ϕ	R
white collar	0.01426	0.041634	0.968965
services	0.03113	0.076898	0.970206
health	0.019721	0.059996	0.968994
retail and food	0.030125	0.078511	0.969769

blue collar	0.030276	0.072824	0.970314