STATE OF CALIFORNIA • DEPARTMENT OF TRANSPORTATION

TECHNICAL REPORT DOCUMENTATION PAGE

TR0003 (REV 10/98)

1. REPORT NUMBER	2. GOVERNMENT ASSOCIATION NUMBER	3. RECIPIENT'S CATALOG NUMBER
CA08-1734		

4. TITLE AND SUBTITLE	5. REPORT DATE
California PECAS Model Development	12/01/2009
	6. PERFORMING ORGANIZATION CODE

7. AUTHOR	8. PERFORMING ORGANIZATION REPORT NO.
Mike McCoy, Ph.D., ULTRANS, University of California at Davis HBA Specto Incorporated	

9. PERFORMING ORGANIZATION NAME AND ADDRESS	10. WORK UNIT NUMBER
University of California, Davis Institute of Transportation Studies 1850 Research Park Drive, Suite 300 Davis, CA 95618-6153	191
	11. CONTRACT OR GRANT NUMBER 65A0296

12. SPONSORING AGENCY AND ADDRESS	13. TYPE OF REPORT AND PERIOD COVERED
California Department of Transportation Division of Research and Innovation, Mail Stop 83 1227 O Street, 5th Floor Sacramento, CA 95814	Final: June 2008 thru June 2009
	14. SPONSORING AGENCY CODE

15. SUPPLEMENTARY NOTES

16. ABSTRACT

A Statewide Integrated Land Use and Transportation Model for California, which is a "prototype operating" version of a Production, Exchange, Consumption, Allocation System (PECAS). The statewide model integrates transportation, economic, demographic, and land use information to predict the exchange of goods and services across space and time.

The development of a Statewide Integrated Land Use and Transportation Model for California. This model will help to understand if integrated models might improve the ability of Caltrans HQ and district staff to better address their goals for congestion reduction, environmental justice and the promotion of livable communities. In addition, the set-up model will assist in understanding how other State agencies might pursue their mandates such as affordable housing and environmental quality, by using an integrated model. This is a functional version of a set-up prototype version of a Production, Exchange, and Consumption Allocation System (PECAS) model depicting future land use scenarios based on simulated micro-economic factors. It will have customization features to include appropriate representation of transportation demand-supply interaction, and congested conditions on relevant transportation modes at relevant locations.

17. KEY WORDS	18. DISTRIBUTION STATEMENT
PECAS, land use, transportation, model	No restrictions.

19. SECURITY CLASSIFICATION *(of this report)*	20. NUMBER OF PAGES	21. COST OF REPORT CHARGED
Unclassified	49	

California PECAS Model Development

Final Report

Executive Summary

Submitted to CalTrans

ULTRANS, University of California at Davis
HBA Specto Incorporated
December 2009

Table of Contents

1. Introduction

1.1. Context (MCM)

- background to project; reason for project
- sponsor
- configuration of team doing work
- specific objectives of project

1.2. This Document

This document describes the work done and the resulting model system obtained in the development of a demonstration version of a PECAS model of California. It sets out the form of the model, the work done in development, the results of base case and policy test runs of the model system, and some conclusions about the success of the work and the appropriate directions of future work on the development of a production version model.

There are 6 sections, including this introduction as Section 1. Section 2 describes the PECAS theoretical framework and the general nature of the design and calibration work done in the development of a specific model using the PECAS framework. Section 3 presents the particulars of the demonstration version California PECAS model, including the specific design, the data sources and how they were used, and the calibration process. Sections 4 and 5 describe the use of the model in a base case scenario run and a demonstration policy test scenario run, respectively. Section 6 offers conclusions.

Sections 2 and 3 can be skipped if the reader is only interested in a more summary indication of the capabilities of the model system and the conclusions arising from the work.

2. PECAS General Framework and Model Development

This section describes the PECAS system in general, as it has been applied in the United States and elsewhere.

2.1. Overview of PECAS

PECAS is a generalized approach for simulating spatial economic systems. It is designed to provide a simulation of the land use component of land use transport interactive modeling systems.

PECAS stands for Production, Exchange, and Consumption Allocation System. Overall, it uses an aggregate, equilibrium structure with separate flows of exchanges (including goods, services, labor and space) going from production to consumption based on variable technical coefficients and market clearing with exchange prices. It provides an integrated representation of spatially distinct markets for the full range of exchanges, with the transport system and the development of space represented in more detail with specific treatments.

Flows of exchanges from production to exchange zones and from exchange zones to consumption are allocated using nested logit models according to exchange prices and transport generalized costs (expressed as transport utilities with negative signs). These flows are converted to transport demands that are loaded to transport networks in order to determine congested travel utilities. Exchange prices determined for space inform the calculation of changes in space thereby simulating developer actions. Developer actions are represented at the level of individual land parcels or grid cells using a microsimulation treatment. The system is run for each year being simulated, with the travel utilities and changes in space for one year influencing the flows of exchanges in the next year.

2.2. Basic Model System Modules

PECAS includes two basic modules that are linked together with two other basic modules to provide a representation of the complete spatial economic system.

The set of four basic modules includes:

- Space Development module (SD module): This is one of the two PECAS modules. It represents the actions of developers in the provision of different types of developed space where activities can locate, including the new development, demolition and re-development that occurs from one point in time to the next. This developed space is typically floor space of various types and is called “space” in the PECAS framework.
- Activity Allocation module (AA module): This is the other of the two PECAS modules. It represents how activities locate within the space provided by developers and how these activities interact with each other at a given point in time.
- Transport Model (TR module): This is one of the “non-PECAS” modules. It represents the transport system connecting locations, including at a minimum a transport network, the transport demands that load onto this network (as a result of the economic interactions represented in the AA module) and the congested times and costs for interactions between locations arising with the loading of these demands.
- Economic Demographic Aggregate Forecasting Model (ED module): This is the other of the “non-PECAS” modules: It is some form of model or approach used to develop aggregate economic forecasts for the study area being modeled. Typically, these forecasts include projected numbers of households or population by category and employment by type (as indications of expected economic activity) for specific points of time in the future. This model or approach may be able to adjust its forecasts in response to information from the AA and SD modules – as is represented in the descriptions included here – or it may provide a static set of forecasts. It may even be the case that there is no model per se that is available, merely the forecast values for the study area. It is also possible to use an extended form of the PECAS AA module to develop such aggregate forecasts – by making some specific assumptions about the relative contributions

to the study-area economy from inside and outside the study area. For the descriptions included here, all of these possibilities are included in the single “ED module” designation that is used.

The four basic modules listed above are linked together with information flows as shown in Figure 2.1.

This linked system works through time in a series of discrete, fixed steps from one point in time to the next, with the AA module running at each point in time and the SD module considering the period from each point in time to the next. In general, the fixed steps can be of any duration, but one-year time steps are recommended since they allow an appropriately quick response of land developers in the SD module to the space prices established in the AA module.

Ideally, the transport model (TR module) used to calculate the congested travel times and associated transport utilities is run for each year, after the AA module has been run for that year. If the overall model run times are too long and travel conditions are relatively stable, the TR module can be run less often to save computation time.

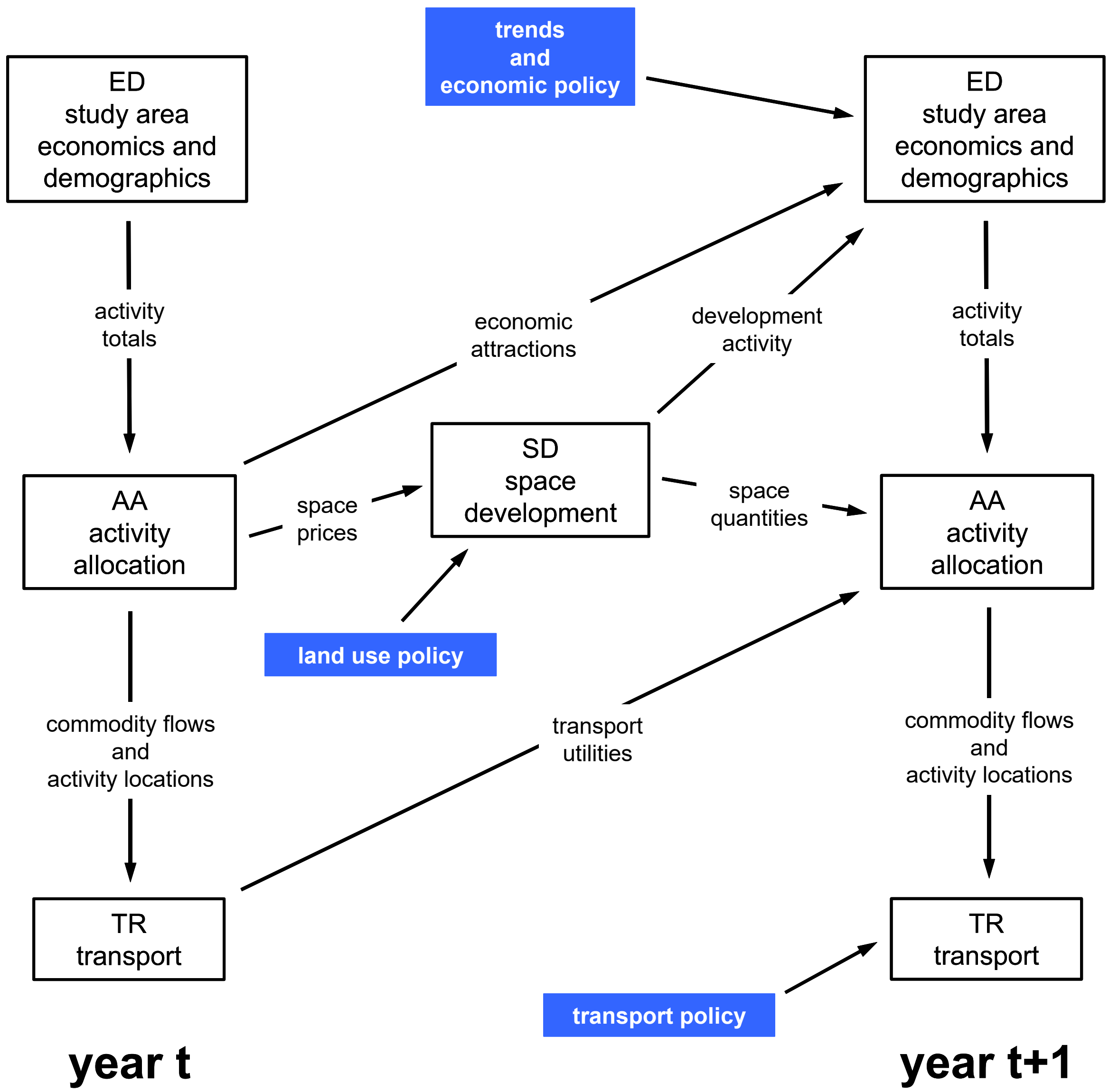

Figure 2.1: Modules and information flows simulating temporal dynamics

The study area is organized into a set of land use zones (LUZs). In the AA module activities locate in these zones and commodities flow between them. Ideally these zones match the transport zones (TAZs) used in the TR module or are aggregations of whole numbers of adjacent TAZs. The connectivity among the LUZs is based on the representation provided by the TR module, where the TR module establishes congested network times and costs and associated transport utilities that the AA module uses in its consideration of the interactions between the LUZs in the next time period.

The land in each LUZ is further partitioned into smaller cells or parcels. The parcels can correspond to actual legal parcels or portions of legal parcels. The cells can be formed by superimposing a grid pattern over the land. The term "parcel" is used to refer to both cells and parcels in the descriptions below. In the microsimulation version of the SD module, developed space (called "space") is located on these parcels, with only one type of space on a given parcel, and the total quantity of each type of space in the LUZs is the sum of the quantities on the parcels in the LUZs.

When an activity in the AA module is located in a zone, it consumes space in the zone, at rates consistent with the production technology or technologies it is using in the zone. Land is used in the provision of the space in the zone, as an input to the development process as represented in the SD module.

2.3. Activity Allocation module

The Activity Allocation module (AA module) is an aggregate representation. It concerns quantities of activities, flows of commodities and markets with aggregate demands and supplies and exchange prices.

Activities are located in LUZs. Activities produce commodities and then transport and sell these commodities; they also consume commodities after buying them and transporting them. There are different types of activities, including industrial sectors, government and households. Activity quantities can be measured in values (for example, dollars of business repair industrial activity) or numbers (for example, number of households with high income and 2 or less persons). The AA module allocates the study-area wide quantity of each activity among the LUZs as part of its allocation process.

Commodities flow at specific rates from where they are produced to where they are exchanged (from seller to buyer), and then from where they are exchanged to where they are consumed. Commodities are grouped into categories, including different types of goods and services, labor and space. Commodities other than space in general flow across zone boundaries. Space is restricted in that it is ”non-transportable” and must be exchanged and consumed in the LUZ where it is produced – which means that the space commodity categories receive some special additional treatments in PECAS as described further below. Commodity flows are measured in values per unit time (for example, dollars of management services per year) or numbers per unit time (for example, tons of coal per month). The movement of these flows of commodities from where they are produced to where they are consumed is the economic basis for travel and transport in the modeling system. It is the travel conditions – the distances, costs, times and associated (dis)utilities by mode – for the movement of these commodities that results in the influence of the transportation system on the interactions among activities and the attractiveness of locations for activities. The AA module allocates the flows of commodities from production location LUZ to exchange location LUZ and from exchange location LUZ to consumption location LUZ, and finds the corresponding set of prices at the exchange location LUZ that clears all markets, as part of its allocation process.

Activities produce commodities and consume commodities in the production process according to the technology they use. More specifically, an activity quantity in a given LUZ produces commodities at specific rates per unit of activity and consumes commodities at specific rates per unit of activity according to the technology being used by the activity. One or more ”technology option” alternatives are defined for a given activity. Each of these technology options is a specific vector of production and consumption rates for different commodities per unit of the activity, representing a particular technology option for the production process available to the activity. The AA module allocates the quantity of the activity in each LUZ among these “technology options” as part of its allocation process.

The allocation process in the AA module uses a three-level nested logit model with a nesting structure as shown in Figure 2.2.

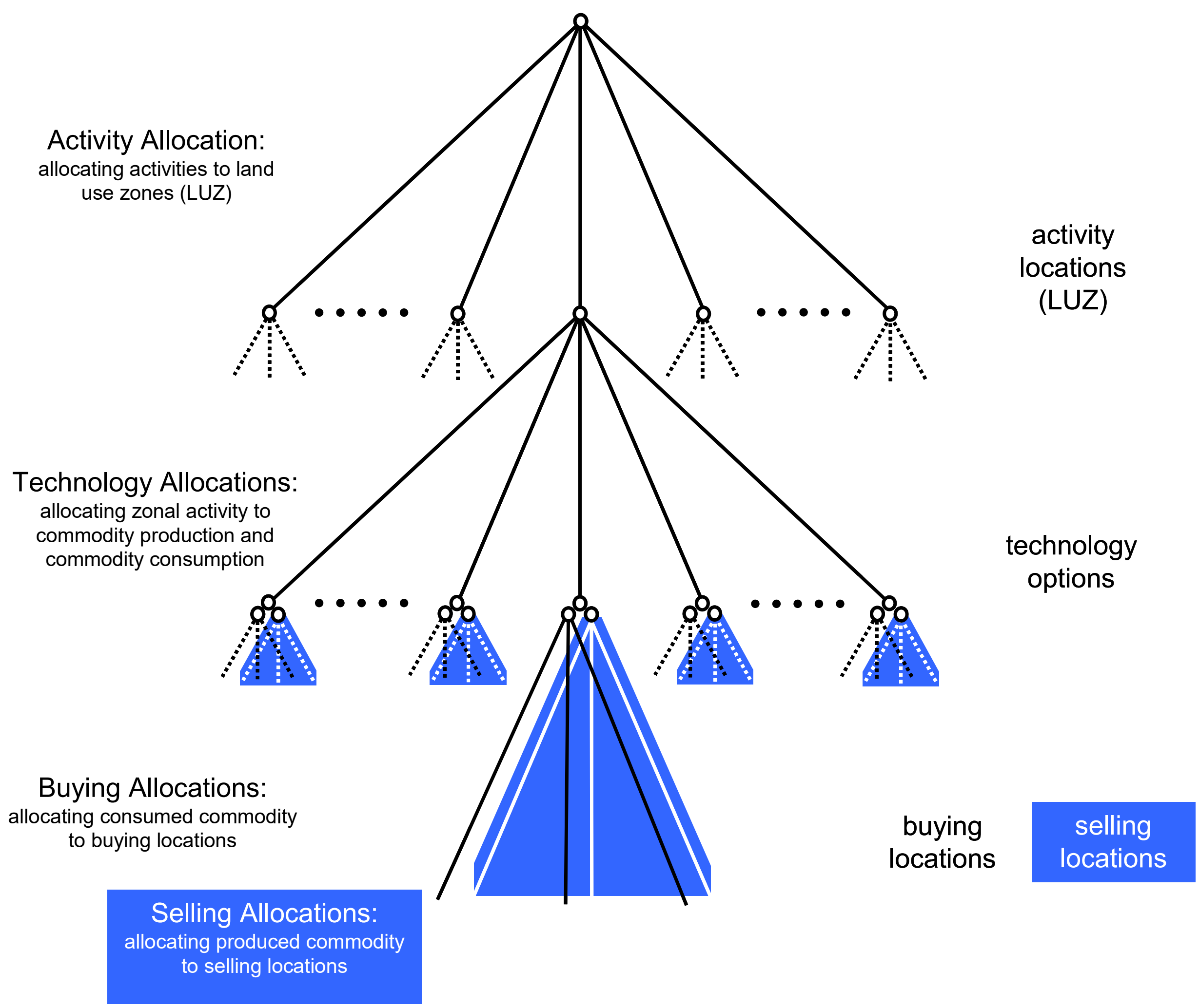

Figure 2.2: Three-level nesting structure used in AA module allocations

At the highest level of the nesting structure, the study-area total quantity of each activity is allocated among the LUZs. At the middle level, the quantity of each activity in each LUZ is allocated among the available technology options. At the lowest level, there are two logit allocations for each commodity in each LUZ: The first is an allocation of the produced quantities among the various exchange locations where they are sold to other activities; the second is an allocation of the consumed quantities among the various exchange locations where they are bought by other activities.

At the lowest level, the utility of each exchange location alternative is influenced by the price at the exchange location and the characteristics for transporting the commodity to or from the exchange location. The composite utility values from these two lowest-level logit models are called the “buying utility” and the ”selling utility” for the commodity in the LUZs. They are used as the transportation-related inputs in the middle-level for allocating the activities in the LUZs among the relevant technology options. The composite utility value for the range of technology options considered at the middle-level for an activity in an LUZ is part of the location utilities used at the highest-level.

The spatial aspects of the AA module allocation process are illustrated in Figure 2.3. Buying and selling allocations link through the exchange locations to establish commodity flows from production to consumption locations in the LUZs.

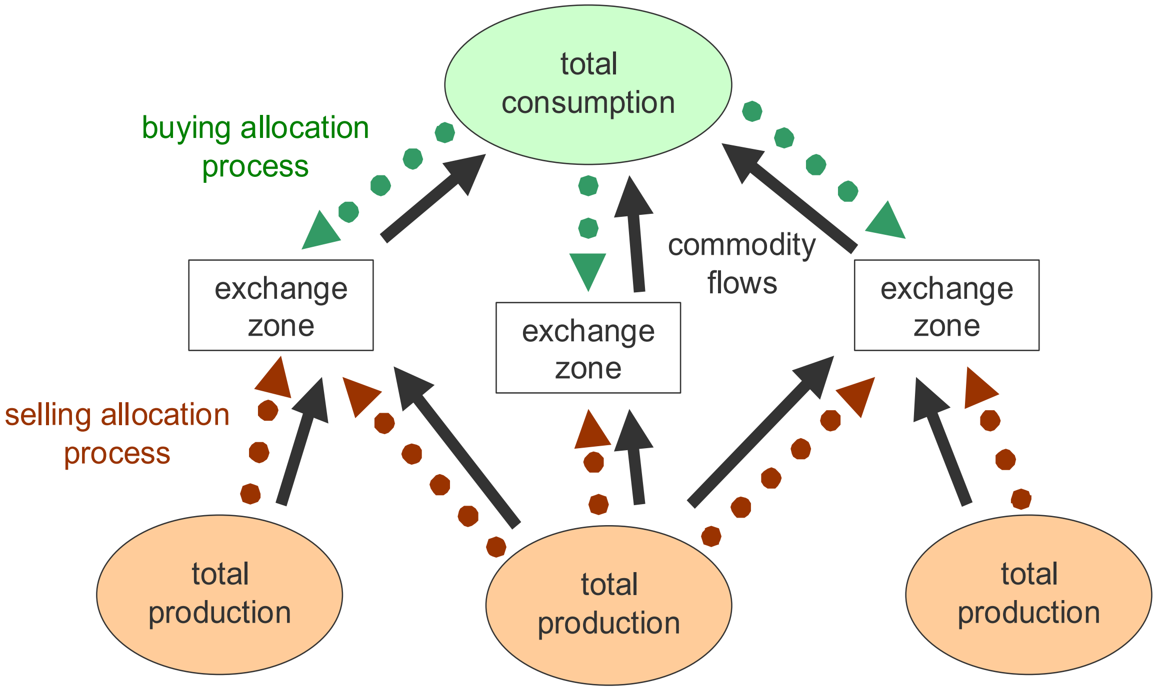

Figure 2.3: Buying and selling allocations resulting in commodity flows from production zone to consumption zone via exchange location

The exchange locations are location-specific markets for commodities, where sellers sell commodities to buyers. Prices are established at exchange locations so that the quantity bought equals the quantity sold – thus the spatial allocation procedure in the AA module assumes a short-run market equilibrium in commodities.

2.4. Space Development module

The Space Development module (SD module) uses a disaggregate approach. It works through a list of the cells or parcels of land in each LUZ, considering each cell or parcel one after another, in each year-to-year step that the module is run.

Each cell or parcel (called "parcel") of land has a set of attributes, including among other items:

- a quantity of existing developed space (called "space") of one type with a specific age;
- a set of zoning rules specifying the types of space that are permitted and the densities at which they are permitted;
- the costs and fees associated with development of each permitted space type and quantity; and
- the price (rent) for each type of permitted space.

Quantities of space on parcels are measured in areas, such as square feet or square meters. Space of a given type is a commodity consumed by activities in the AA module. Space is "non-transportable" in that it must be consumed in the zone where it is located. The aggregate quantity of space of a given type in a given zone is the sum of the quantities on all of the parcels in the zone.

Developers act to change the types of space and/or the quantities of space on parcels. The consideration of the cell or parcel in a given year-to-year step includes determining the development event for the year-to-year step, establishing whether or not the existing space is changed by some form of developer actions during the step, and, if the space is changed, what the updated space type and quantity are to be. Keeping the space the same in a given year-to-year step is also a development event in this context, as is allowing the space to go derelict through neglect.

In a given year-to-year step, a Monte Carlo process is used in the determination of the developer event for each parcel. Logit models are used to assign selection probabilities to each of the events that are permitted for the parcel according to the zoning rules, including each of the permitted updated space types and quantities options, along with the "no change" and "derelict" events. One of these events is then selected and results recorded, and the process moves on to the next parcel. The utility functions in these logit models calculate the expected net revenues to the developer for the available options, incorporating the prices and the costs for transition, maintenance and servicing in each case.

The space prices that are considered are expressed per unit of space per unit of time considered in the AA module, and thus are rents for the use of the space for a given time and not prices for purchase of the space or the associated land for ownership in perpetuity.

2.5. PECAS Model Design

2.5.1. Nature of design in this context

The basic functionality of the PECAS framework is fixed. The flexibility is in the categories and functional forms and parameters used within the basic functionality of the framework.

The design of a specific PECAS model in this context is the identification of the categories for the model components and the functional forms for the interactions among these components.

The categories and types to be defined include:

- activity categories, with groupings of
    - economic industry sectors
    - government

    - production factors
    - households
- commodity categories, with groupings of
    - goods
    - services
    - financial flows
    - labor types
    - space types
- exchange types for commodities;
- land categories;
- land use zones, as aggregations of transport analysis zones;
- import-export zones;
- time steps between runs of the TR module;
- land zoning schemes; and
- linkages between commodities in the AA module and travel segments (or purposes) in the TR module.

The functional forms for the interactions among these components are selected from those available at each point. The values for the parameters in these functions are determined in calibration.

2.5.2. Approach in design

The guiding objectives in the design of a particular model, in order from most to least important, are as follows:

- provide appropriate policy sensitivity;
- establish complete coverage before detailed representation; and
- provide detail according to analysis needs, not data availability or unavailability;
- start simple and then add detail and complexity in development iterations; and
- take into account data unavailability but do not overreact to it, and thus be prepared to synthesize data according to theory if required.

Providing for appropriate policy sensitivity requires sufficient disaggregation in the range of relevant categories. For example, if alternatives are to be considered regarding public housing provision and eligibility, then public and private residential housing should be represented as separate space categories and eligibility should be used as a basis for defining household categories in order to provide appropriate representation of the responses and resulting impacts.

The identification of the policy options to be considered with a model can sometimes be an appropriate formal task that is performed as part of the model design work. This can include a review of relevant planning documents and discussions with policy-makers and analysts with knowledge and experience in the area to be modeled. Specific aspects of the policy options that can help inform the model design include the nature and the timing and geographic scope of the proposed differences among options and among the anticipated impacts of these options. The appropriate levels of disaggregation and detail in different parts of the model are influenced by these aspects.

Design considerations draw extensively on previous experience in this sort of model development and application work gained with this theoretical framework, along with its forerunners and with other related frameworks.

2.5.3. Design Diagram

A structured diagram, called a PECAS “design diagram”, is used to depict key aspects of a model design. This diagram shows the defined categories for the model elements and the treatments of their interactions.

The basic structure of a design diagram is shown in Figure 2.4.

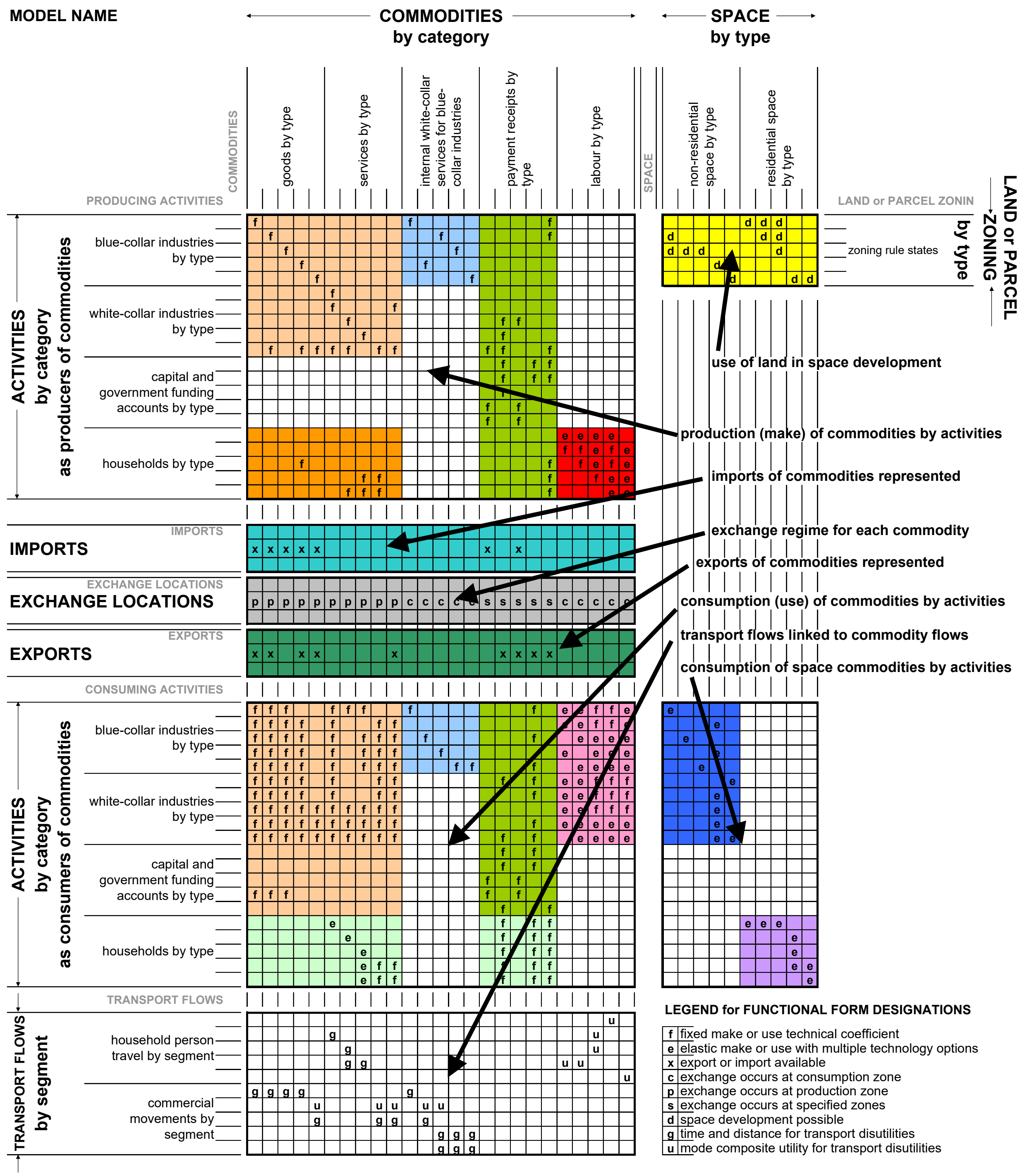

Figure 2.4: Design diagram format indicating contents of component tables

The design diagram is a set of aligned matrices that together indicate the categories of the different elements included in the model and the functional forms used to represent the interactions among these elements. In each matrix in the diagram, the columns represent categories of commodities (with types of space included as commodities) and the matrix itself concerns some aspect of the treatment of these commodities.

The different background colors within each matrix indicate different components of the treatment represented by the matrix – these components defined (as shown by the colors) according to shared aspects of the approaches and data needs in model development. These different "color components" are used to structure the discussions about data needs and data processing included in section 6 in this document.

The single-letter characters in the cells of the matrices indicate aspects of the representation of the interaction between the corresponding row and column items. For example, the "f" in the upper left-hand corner of the upper left-hand matrix indicates that the activity listed in the first row (in this case a category of blue-collar industries) produces the commodity listed in the first column (in this case a category of goods) at a fixed rate. The definitions of these single-character codes are listed in the lower right-hand corner of Figure 2.4.

2.6. Agile Approach to Model Development

A particular PECAS model application is developed in a series of "development iterations". In each of these development iterations a version of the complete model providing coverage of the geography, time and set of actors is established and documented, building on what has been established in previous iterations. Early iterations use very aggregate and simple representations for components as required in order to avoid longer durations than a few months. The first iteration considers a version of the AA module with perhaps five zones, five activity categories and five commodity types along with a rudimentary placeholder for the SD module; it is as much about load testing the computer setup and putting in place the documentation systems as it is about the representation provided.

This use of development iterations is in stark contrast to a more traditional line process where each model element is assembled once to final specification on its own and then the elements are combined in order to realize the full model system. In early development iterations after the first, the focus is on the more risky elements of the work, where there is comparatively more uncertainty about the ability of the resulting model to provide a representation as required. This is to avoid serious failures of design or approach emerging late in the project when they are more difficult to correct.

The calibration effort in each iteration includes adjusting the model mathematical functions and coefficient values in order to improve the model's representation of the real world. A full PECAS model is very complex and intertwined, like the real world the model is seeking to represent. Calibration to finer and finer levels of tolerance across the full range of model dimensions can take a very long time, much longer than the appropriate time for a development iteration. Practically speaking, some sort of compromise on the degree of calibration is often required in a given iteration, with further improvements left to later iterations. A "quick and dirty" calibration may be the most appropriate in some instances.

2.7. Calibration

The calibration activity in each development iteration includes the adjustment of model mathematical functions and coefficient values in order to improve the fit of the model to the real world.

In general, a three-stage process is used in calibration. This three-stage process is illustrated in Figure 2.5 and described below.

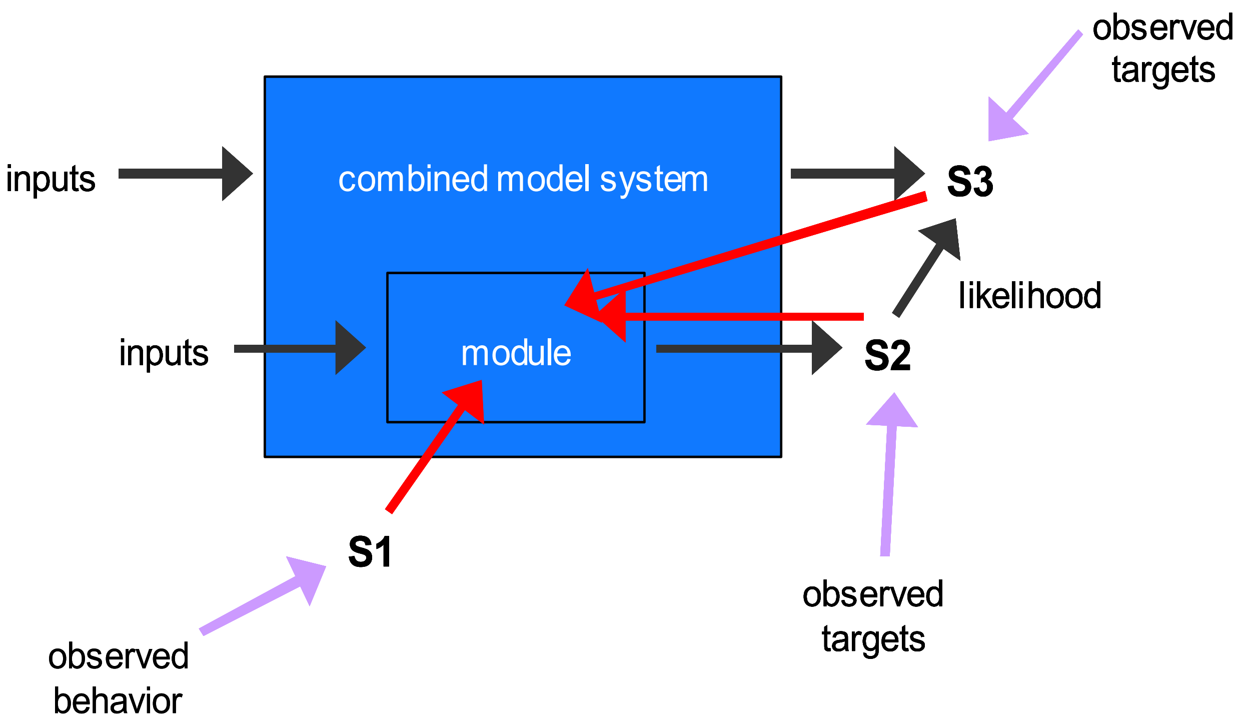

Figure 2.5: Three-stage process used in calibration. The values for parameters are established in a three-stage process. In the first stage, the values for the "S1" parameters are determined using external data and statistical methods and remain largely fixed through the rest of the calibration. In the second stage, the values for the "S2" parameters for each module are determined using the fit of the module against targets. In the third stage, the values for the "S3" parameters, which are a selected subset of the "S2" parameters, are determined jointly for all modules using the fit of all modules against targets.

In the first stage, values are established for certain "S1" parameters in each of the AA and SD modules, along with those in the ED and/or TR modules to the extent that these other modules are also being developed in the project. It is unlikely that these values established for these S1 parameters will be adjusted as the model development and calibration work progresses in the current development iteration, although it is not impossible that some may be adjusted. But these values will be superseded by other values arising in subsequent development iterations, at least to the extent that these values relate to components of the model that will be further disaggregated or otherwise redefined in such development iterations. At this point it is not necessary that the entire SD or AA module can be run; the components of the module in each case are being "assembled" and the outputs of the module are not yet being considered.

In the second stage, initial values are established for all of the parameters that are not S1 parameters, called the "S2" parameters, in each module, considering the aggregate fit of each of the AA and SD modules in isolation. This aggregate fit concerns specified targets for outputs from the module, so the module needs to be run on its own in order for it to provide these outputs. Thus, a full set of required inputs for the module needs to be developed, including all those provided by other modules and all those provided exogenously. In order to obtain reasonable values for the S2 parameters, these inputs need to be consistent with the specified targets, representing conditions similar to those that gave rise to the targets.

In the third stage, the initial values established for certain sets of the S2 parameters are revisited for the full model system of all modules simultaneously, considering the fit of all modules together, with the full model running, so that inputs to the modules are coming from the other modules in the way they do in a model run. These revisited S2 parameters are designated "S3" parameters. The second and third stages together constitute a Bayesian updating process for these S3 parameters. Ideally, all parameter values would be revisited, but this is not possible for practical reasons. A weight sensitivity matrix can be used to explore the remaining lack-of-fit for the entire model, which can help identify the parameters to focus on in the third stage and which may lead to comparatively minor changes in the details of the model design and specification.

3. California PECAS

3.1. Design

3.1.1. Intentions

The goal of model design is to determine the activities, commodities, and space types in PECAS and land use types in county and city general plans on condition of the constraints in the client's need, budget, staff, deadlines for deliverables, data availability and computation facilities. The general principle of model design is that the activities (households, blue-collar industries, service industries, and government accounts) and commodities (goods, services, labor, tax receipts, household investments, money transfer, space type) should well represent the structure of economy and the land use in California, while considering the constraints and limitations of expected data and the expected policies use of the model.

3.1.2. Design diagram and its interpretation

According the general rule of design stated above, we identified 65 activities (48 blue-collar and white-collar industries, 3 financial accounts and 14 household types) and 85 commodities (49 types of goods and services, 3 financial flow commodities, 19 labor types, and 14 space types). The model design is represented by a matrix called a "design diagram", as described in section 2.5.3. The design diagram for the demonstration California PECAS model is attached as Appendix 1. In the design diagram, the first column lists production activities, consumption activities, exports, imports, and exchange locations, and the first row lists the commodities. The MAKE of commodities is represented in the top half of the diagram, while the USE of commodities by industries is represented in the bottom half. The blue-collar and white-collar industries consume (USE) goods, services, and labor at a fixed rate (which is represented by f in the cells) and space at an elastic rate (which is represented by e in the cells) to produce (MAKE) goods and services. Households produce labor at an elastic rate and consume goods, services at a fixed rate and space at an elastic rate.

3.1.3. Calibration year and calibration period

In the Demonstration model, the calibration year is 2000, to allow maximum use of year 2000 census data, while the calibration period spans from 2001 to 2006.

3.1.4. Zone systems

In the Demonstration model, we use two zone systems in PECAS: The Transportation Analysis Zone (TAZ) system and Land Use Zone (LUZ) system. The TAZ system has 4683 TAZs, and is the same as that in the high speed rail (HSR) statewide travel model. The purpose of adopting this system is to keep the number of households and workers in PECAS consistent with those in the travel model in the base year so that the outputs from PECAS can be easily integrated with the travel model.

The LUZs are the spatial units in Activity Allocation module (AA) of PECAS. A LUZ is a cluster of spatially continuous TAZs. All the results of AA are reported at LUZ level. In the Spatial Development (SD) module of PECAS, the spatial unit for the choices made by land developers is 50 meter grid cell, but the results are summarized and reported by TAZ and by LUZ. In the Demonstration model, there are 523 LUZs. The number of LUZ in each county is listed in Appendix 2.

3.2. Data Used in Development

This section describes the data used in the construction of the Demonstration version of the California PECAS model. It describes the data in terms of the representation of behavior in the model.

The plan for the data collection for all phases of model development was described in Technical Memo 2: “Data requirements and estimated costs for the full zonal demonstration model” which has previously been submitted to CalTrans. This technical memo identified and numbered model elements in a comprehensive manner facilitating the identification of necessary and desired data.

3.2.1. Review of model element numbering system from Technical Memo 2

The Activity Allocation (AA) module represents three basic choices for the Activities (employment and households) in California:

1. The choice of "exchange location", being the choice of where and who to interact with while exchanging goods, services, labor or space. The utility function for this choice contains four parts: Transport Costs (element **2.1.1**), Prices (element **2.1.2**), Size (element **2.1.3**) and Variability (element **2.1.4**);
2. The choice of "technology", being the quantities of different categories of goods, services, labor or space to consume or produce. This is sometimes called "lifestyle" in the case of households, where it is element **2.2.1**. In the case of employment activity (goods and services production by business or government) it is element **2.2.2**;
3. The choice of home location, which is element **2.3.1** in the case of households, or element **2.3.2** in the case of employment activity.

AA views the rest of the world through its representation of import and exports, which is element **2.4.** The quantity of space available in any one year in AA handled by a representation of landlord decisions, element **2.5**, called the "short term floor space supply function". This is separate from the inventory and construction of space between years, which is represented in the SD module.

In the Space Development (SD) model, the decisions of the building construction industry are represented with three choices:

1. The choice of the intensity of development, i.e. the total floor space in the new building divided by the amount of land holding the building. The utility function for intensity choice contains construction cost (element **3.1.1**), local effect rent modifiers (element **3.1.2**) and baseline prices (included already in element 2.1.2), random variation (element **3.1.3**), permitted intensities as a policy input (element **3.1.4**), and building maintenance costs (element **3.1.5**);

2. The choice of the type of building to construct, which contains, in addition to elements 3.1.1 through 3.1.5, the space transition constants (element **3.2.1**) and the impact of zoning which disallows certain developments (element **3.2.2**);
3. Whether to build a building in a given year on each grid, which is largely determined from elements already listed but also includes a single parameter controlling the rate of development (element **3.3**).

Overall changes in the economy over time (statewide control totals) are specified as changes in quantities of households by type (element **4.1**), changes in industry size and mix (element **4.2**), response of the rest-of-the-world to changes in California (element **4.3**), and changes in technology (element **4.4**).

Finally, in the development of the starting GIS grid database for the base year, we require total amounts of floor space by our PECAS coding system for each LUZ (element **5.1**) and spatial (GIS) information describing the land in California (element **5.2**).

3.2.2. Data used in the development of the Demonstration model

This section describes the data used in the development of the Demonstration model, its sources, and for which model element it was used.

A. IMPLAN model

The IMPLAN model is an economic model of the financial relationships between various sectors of the economy. It is our starting point to determine the relationships that we need to model spatially in PECAS. Since money changes hands constantly in society, we can understand many relationships and interactions by understanding the flow of money from seller to buyer. In the demonstration model IMPLAN relationships were used for model elements 2.2.1, 2.2.2, 2.2.3, 2.4 and 2.5.

B. PUMS (Public Use Microsample) data from census

The census sample of individual responses provides a detailed record of individual lifestyle choices by households, and limited information on the jobs performed by household members. These data were used for model elements 2.1.3, 2.2.1, 2.2.2, 2.3.1 and 5.1.

C. Census Summary File 3

The Census Summary File 3 provides spatial information on the types of households living in different locations. This was used in the development of model elements 2.1.3, 2.3.1 and 5.1. Census self-reported home value data and home rental cost data were used in elements 2.1.2 and 2.3.1.

D. Elevator permit data

A limitation of the census data is that large multifamily buildings are identified based on the number of units in the structure, thus high-rise buildings (which use land very intensely) are not distinguishable in the census from horizontally large buildings, such as sprawling two and three story garden apartments. The database of elevator permits in California was acquired to identify where the taller buildings are. This was critical in helping to identify the quantity of land covered by residential structures in areas with multifamily dwellings, in elements 2.1.3, 2.3.1, 5.1 and 5.2.

E. Travel model times and costs

The California High Speed Rail travel demand forecasting model was used to estimate zone-to-zone travel times and distances, for use in element 2.1.1. The High Speed Rail model did not have appropriate responses in some aspects that were not closely related to high speed rail, so a factoring process was developed to adjust the High Speed Rail Model’s responses.

F. Real estate data on prices for residential and non-residential space

Real estate professionals (Grubb & Ellis Research, Colliers International, Cornish & Carey, CB Richard Ellis) were contacted and they provided publications showing the current average price of different types of real estate in different market areas. These were compiled into a geographic coverage layer of space prices, and used in element 2.1.2 and 5.1.

G. InfoUSA data on location of employment

A private company, InfoUSA, sells data on the location of business establishments. This was processed and cleaned up and compared with the IMPLAN model's employment counts. These data were used in the development of elements 2.1.3, 2.3.2 and 5.1.

H. Oregon and other data on space use rates per employee

For the demonstration version of the model, the amount of space used at work for each employee was determined from functions developed for other PECAS models, notably the Sacramento model and the Oregon model. These were used in elements 2.1.3, 2.2.2 and 2.3.2.

I. Planned land use

Long-term land use regulations are a policy input to the model, to be modified as part of scenario exploration. However, a base case scenario must be developed that best reflects current knowledge regarding plans for future land use regulations. The general plans of California were compiled and processed into a consistent representation of future regulations. The regulation categories were interpreted in terms of the type of construction that is allowed and its density; these interpretations are for demonstration purposes used in elements 2.1.3, 3.1.4 and 3.2.2.

J. Construction cost data

Initially, the functions predicting the costs of construction were transferred from other PECAS models. The RSMeans construction cost database was then purchased to further refine these functions for California conditions. These were used in element 3.1.1.

K. Forecast total quantity of floor space

Region wide space construction rates growth rates were set to be 2.4% per year. For calibration, these were then disaggregated using urban-vs-rural profiles. This disaggregation was based primarily on expert experience. These were used in element 3.3 (and are discussed further in section 3.3.4 of this document).

L. Goods movement shipping cost data

The costs of shipping freight were developed from the 2002 Commodity Flow Survey Table, using truck average miles. These were adjusted to separate out time effects (primarily using truck driver's wage information) from distance effects, so that freight decisions respond to congestion; these were used in element 2.1.1.

M. Services movement travel cost data

The costs of delivering (or picking up) services were developed from the 2000-2001 California Statewide Travel Survey Weekday Travel Report, and used in element 2.1.1.

N. Personal trip generalized cost data

The personal burden of travel as a function of trip type, trip distance and trip time were developed from the coefficients in the California High Speed Rail Travel Model. For work trips, hourly wage rates from the Bureau of Transportation Statistics were used to adjust time values. These were used in element 2.1.1.

O. American Housing Survey

The decennial census only reports dwelling size in terms of number of rooms; the American Housing Survey was needed to convert these rooms into a measure of dwelling sizes in square feet, required for elements 2.1.3, 2.3.1 and 5.1.

P. Growth rates

For the demonstration model, statewide population and employment growth were set at 1.2% per year, used in elements 4.1 and 4.2.

Q. Trip length information

The Commodity Flow Survey was used to determine an appropriate length for goods movement trips. For passenger trip miles, we used the California Statewide Travel Survey and other travel surveys. For labor commuting distances we used the tabulations from the Census Transportation Planning Package (CTTP) regarding the locations of workplaces.

R. Synthetic Base Parcel Database

An arbitrary 50 meter grid was developed to encompass the geographic area of California. Each of these 50 meter square cells was then defined in PECAS as a "parcel". The median housing age of each cell was then extracted from 2000 U.S. Census block groups. The zoning classification for each cell was derived from a generalized general plan GIS dataset previously developed at ULTRANS. The general plans were collected from cities and counties in CA and then combined in GIS and reclassified into 13 categories for use within PECAS.

The grids were then overlaid with the 2000 National Land Cover Dataset, the 2000 U.S. Census block data, the generalized general plan dataset and layers to represent the public land mask. This was element 5.2, which was used with the PECAS space synthesis procedure to assign a base year PECAS space type and PECAS space quantity to each grid cell.

3.3. Stage 2 Calibration

After the model was constructed using the data sources described above, certain data sources were also used in what is called “Stage 2” calibration. As discussed in section 2.7, Stage 2 calibration involves running individual elements of the model and comparing the results from the model run with certain data describing the real world, and adjusting certain parameters in the model so that it best reproduces observed patterns.

In the California Demonstration model there are several procedures that were used in this Stage 2 calibration.

3.3.1. Trip length calibration

The variety of each commodity in the model affects trip lengths in California; certain commodity categories (representing goods, services and labor) are relatively homogeneous, and one would only travel a long distance to buy or sell these things if it were financially rewarding (cost savings exceeded transport costs) or if it were not available nearby. Other commodities are more heterogeneous, and a nearby option may not be appropriate even though it may be well-priced and nearby. We calibrate the degree of heterogeneity in each commodity by adjusting one parameter for each commodity until the distances for trips involving that commodity match observed distances.

These parameters are critically important in PECAS – they control the degree to which large economies, such as California's, are able to prosper due to the wide variety of goods, services and workers on offer. This calibration procedure was automated in the demonstration model so that these critical parameters can be updated as new and better data is available during production model development.

3.3.2. Option size calibration

At the middle level of AA is the choice of how much of different labor occupation categories to produce and consume, and how much space to consume. When we expand the model to allow these choices in future years and future scenarios, we also need to take on the task of adjusting certain parameters so that the structure of the occupation and floor space markets still matches observed base year conditions. An automated process was developed for this calibration since the large number of options in the technology specification make manual calibration difficult. Repeated runs of the AA model can be performed over the course of days, with the automated procedure adjusting parameters until the model represents base year reality even with the flexibility for future year adaptation.

3.3.3. Floor space calibration

Observed floor space totals from the data sources described in section 3.2 will never be consistent with each other in a large model area such as California, and the categories used in observed data will never exactly match the categories in the model design. Thus it is appropriate to use data that is more consistent across the state as the controlling data, and to develop quantities of modeled floor space from the more consistent data. In the California Demonstration model, the population data, the employment data, and the floor space price data were taken as the statewide consistent data sets. These were used together with information on quantities of floor space used by households and employees to establish a quantity of floor space. It is necessary to use the model itself, in repeated runs, because floor space use is elastic with respect to floor space price but in a complex relationship with the other commodities represented in the AA module.

Thus the procedure for establishing floor space quantities in each zone involved first setting up AA so that it respected population and employment numbers, then calibrating AA's space use equations, and then adjusting the quantities of floor space in each LUZ until the price produced by AA for each space type matched the observed data.

An automated script was written to perform this adjustment, and the AA model was able to reproduce floor space prices in the base year.

3.3.4. SD dispersion parameters and transition constants

Data on the quantity of development in California was not directly collected (in appropriate units) for this Demonstration model. However a procedure was developed to enable the calibration of SD to quantities of development. These procedures were tested, and the SD model was roughly calibrated, by factoring assumed construction rates to have different construction rates in different areas of California.

Ten different subregions of California were selected to represent a cross section of different urban, rural, coastal, mountainous and desert areas with different rent values for different space types. The SD module should be able to produce appropriate levels of construction activity statewide, but also appropriately different levels of construction activity in areas with different rent profiles.

Again an automated script was written to run SD repeatedly, adjusting the constants associated with each space type to achieve the correct total amount of construction, and adjusting the "dispersion parameters" within SD to achieve an appropriate difference between the different ten subregions.

This calibration was only possible because the team had overcome some initial computing issues. Initially, the SD module was developed using one database system, but the size of the state required SD to be setup using a different, more robust, database system. The SD software was also adjusted to be substantially faster; much of this worked involved understanding the limitations on speed imposed by the database system itself.

The resulting SD model produces appropriate levels of construction in appropriate zones, as will be shown in section 4.

4. Base Case Scenario

4.1. Purpose

4.1.1. Running the model into the future

Running model into the future involves interactions between AA and SD and between AA and the travel model. By running the model into the future, we want to test 1) if AA converges in each forecast year; 2) if AA and SD interact as expected; 3) if AA and the travel model interact as expected; 4) if the calibrations of AA and SD are appropriate; 5) the time for the integrated model to finish a full run. In the Demonstration model, we run the model from 2000 to 2020 at a one-year time step. AA interacts with the travel model at a five-year time step, i.e., in 2005, 2010, and 2015.

4.1.2. Provide reference case

Besides the goals listed above, another goal is to set up a reference for a sensitivity test of the integrated model.

4.2. Development

4.2.1. Inputs needed for running the model into the future and how they were developed

The inputs of AA include:

- Properties file: This files includes the commands which controls the inputs, outputs, and model convergence. Each year has a properties file.
- ActivityTotalsI: This is an exogenous input table which has a control total for each activity in each year at the model level. In the base year (2000), the activity totals are estimated based on the Census Block Groups SF3 file, the number of households and workers in the travel model, and InfoUSA database. Starting from 2001, the activity totals are calculated at a 1.2% growth rate, which is the average annual growth rate between 1990 and 2000.
- PecasZonesI: This table includes all LUZs and TAZs.

- ActivityConstraintsI: This table includes the activity totals in each TAZ in the base year. The command to use the constraints is in the properties file in year 2000. This command forces AA to adjust zonal constants to generate the same results as in the table. The table is generated by splitting TAZ activity totals with the proportions in the InfoUSA database (for employment) and census (for households).
- TechnologyOptionsI: This table includes technology coefficients for industry/goods MAKE and USE, industry/services MAKE and USE, industry/labor USE, industry/space USE, household/goods MAKE and USE, household/services MAKE and USE, household/labor MAKE, household/space USE, imports, and exports. The data used to calculate the coefficients include IMPLAN social accounting matrix, Census 1% Public Use Microsample (PUMS), and national Commodity Freight Survey.
- ActivitySizeTermsI: This table is used to adjust the weights of activities based on their proportions in the base year. In the Demonstration model, we give the same weight to all the activities.
- CommoditiesI: This table includes the commodity coefficients which draw the demand and supply curves. The unit transport costs by time and distance are calculated separately by using IMPLAN social accounting matrix, commodity flow survey data, and PUMS.
- ExchangeImportExportI: The coefficients in this table define the import and export curves. The coefficients are calibrated based on targets in the base year.
- FloorspaceZonesI: The floor space zones in this table are TAZs.
- FloorspaceSupplyI: The coefficients in this table define the short term space supply curves. The coefficients are calibrated with the space supply relative to the space in the base year.
- DevelopmentTypesI: This is an input table to SD. It lists space development types and corresponding development costs (construction costs, reconstruction

costs, maintenance costs, age related rent discount factor, etc.). We borrowed these factors from the Sacramento Council of Governments (SACOG) model.

- ZoningSchemesI: This table lists land development types allowed in each land use type in the general plan and corresponding development fees.
- TransitionConstantsI: This table lists possible transitions from one land use type to another. The coefficients are calibrated based on the transition rates from 2000 to 2001. See Appendix 3.
- FloorspaceI: This table includes the total quantity of space for each space type. In the base year, it is the space inventory. In the forecast year (2001 to 2020), this table is generated by SD.

4.3. Running the Base Case Scenario

In the Base Case scenario, we set up a fixed annual growth rate for households and employment. In our eight-core server, AA takes about 6 hours in 2000 and 50 minutes per a year from 2001 to 2020. SD takes about 50 minutes to finish a one-year run for the whole state. The travel model takes about 10 hours to finish a run. Thus, the total run time from 2000 to 2020 is about 69 hours. This process is not automated in the Demonstration model but will be automated in the Production model.

4.4. Scenario Results

Figures 4.1 through 4.5 show some results for this base case scenario. They are discussed below.

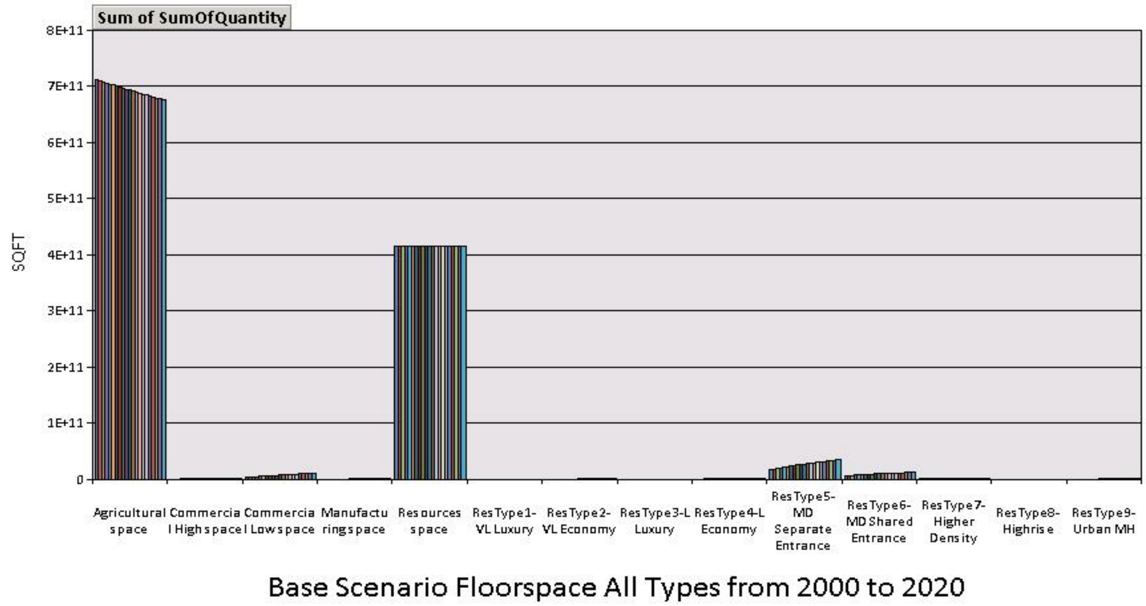

Figure 4.1 Total space from 2000 to 2020 in the Base Case scenario

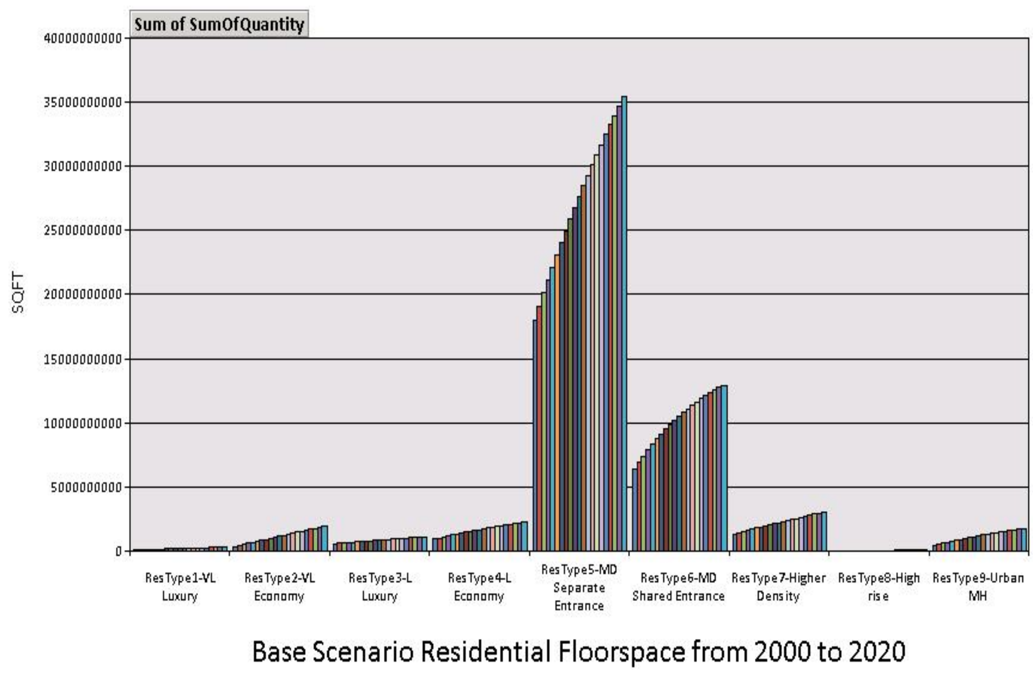

Figure 4.2 Total residential space from 2000 to 2020

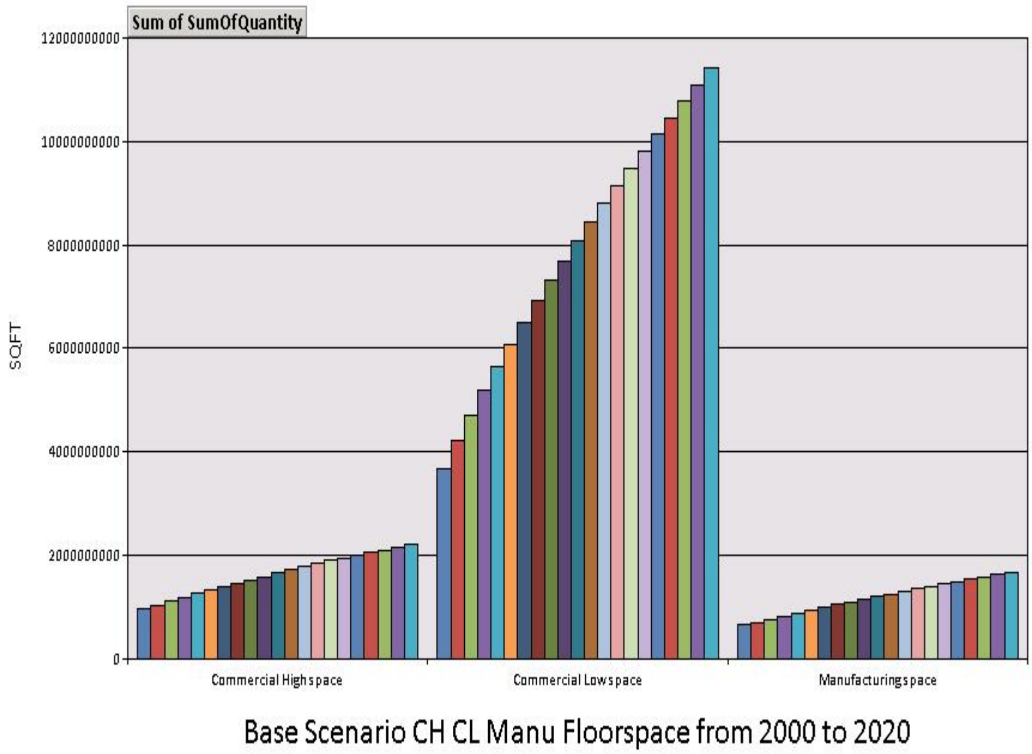

Figure 4.3 Total commercial and manufacturing space from 2000 to 2020

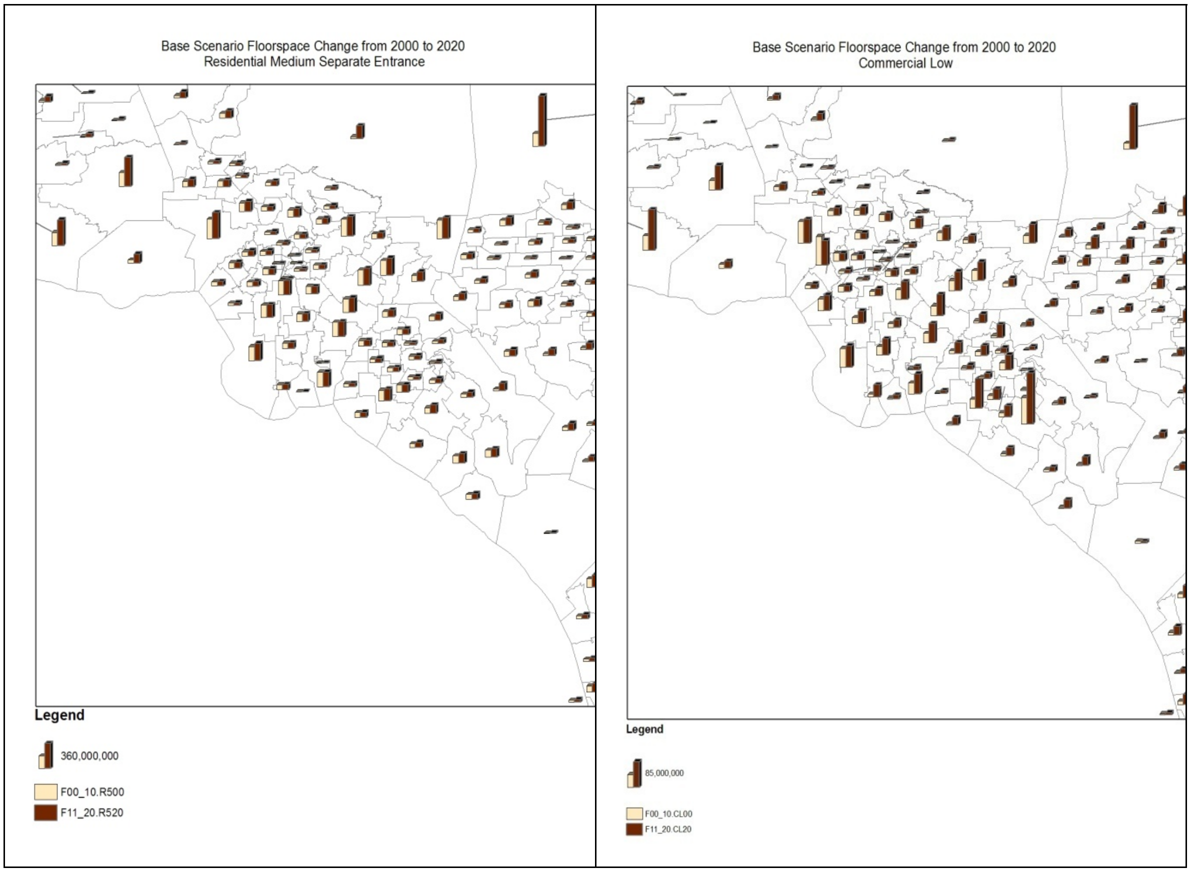

Figure 4.4 Residential medium separate entrance floor space (left figure) and commercial low floor space (right figure) in 2000 (left bar in each figure) and 2020 (right bar in each figure), Base Case scenario, Los Angeles area

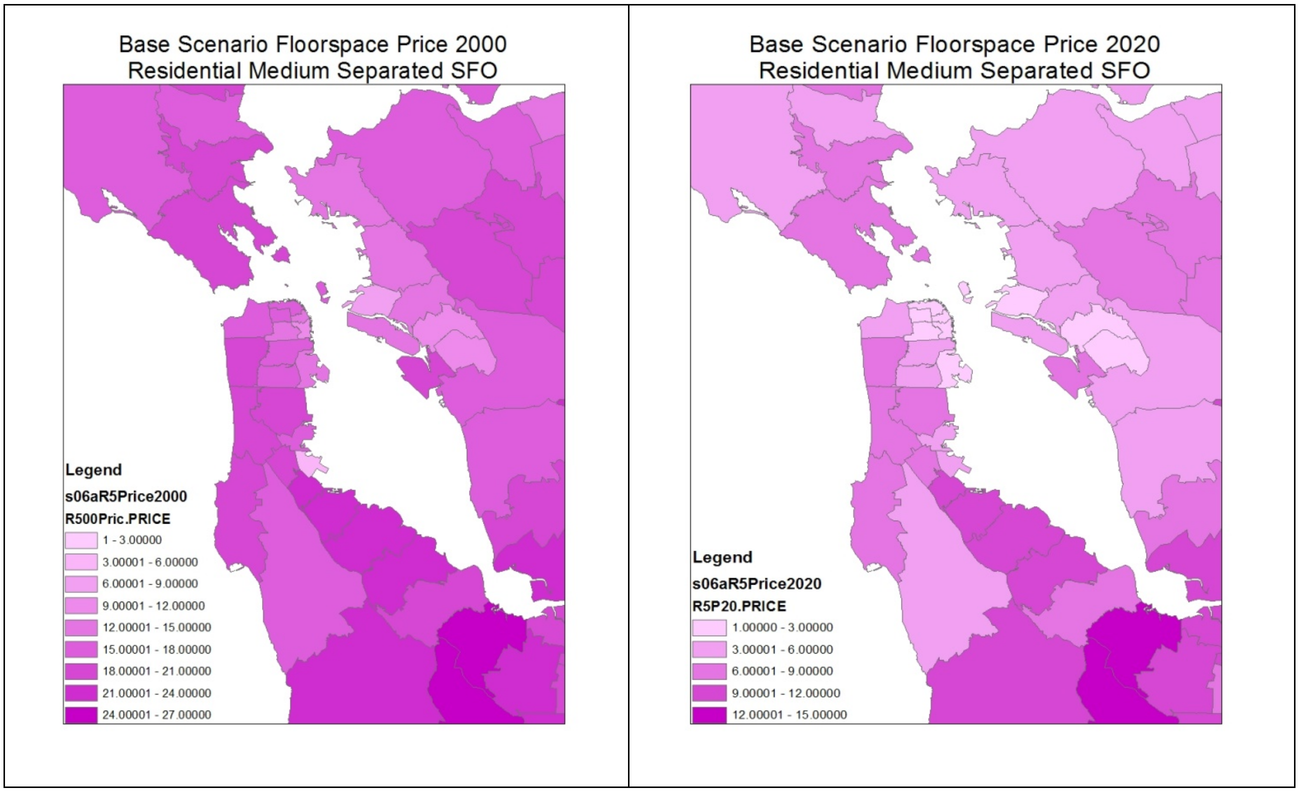

Figure 4.5 Space price (2000 dollars per square foot) for typical single-family detached house in 2000 and 2020, Bay Area

4.4.1. Discussions and findings from the Base Scenario

In the Base Case model, we hypothesize the following model responses:

- Agriculture land area would monotonically decrease over time.
- The trend in development of space in manufacturing, commercial, and residential would be consistent with that of employment and household growth.
- Space price will decline in general due to abundant supply of space.
- In SD, agriculture space is allowed to be converted into all other space types but not vice versa. Not surprisingly, agriculture space decreases over time due to the growth of population and employment (see Figure 4.1). Resources land area

represents military space, and is designated for military bases only. Thus, the quantity does not change over time. The total quantities of other space types grow at different rates over time (see Figures 4.2 and 4.3). Single family separate (ResType5, the dominant space type in California) and shared entrances (ResType6) have higher growth rates. This is plausible because we don't expect a large redistribution amongst different types of housing in a "base case" scenario. Low density commercial uses more space and has a larger growth rate than high density commercial and manufacturing, reflecting the land use designations in general plan and the general usefulness of commercial low density space for a wide variety of production activities,

Figure 4.4 is a snapshot of the spatial development pattern in the Los Angeles area. The residential medium separated entrance has a higher growth rate in suburban or exurban zones than in the inner urban areas. The vacant land and residential space supply in those zones are abundant and make residential space price lower over time. At the same time, the travel cost stayed unchanged. Therefore, the marginal benefits of choosing to live in suburban or exurban areas are higher. The growth pattern of the commercial low-density space shows an opposite trend, i.e., a higher growth rate in the inner urban areas than in the suburban or exurban zones. This pattern can be explained by the agglomeration effects of commercial activities and the relatively small amount of land used by commercial space (when compared to residential space), which enables a larger percentage growth of commercial in existing built-up areas if market conditions are favorable.

In the Base Case scenario, we do not constrain construction activities but calibrate SD to a growth rate of 2.4% from 2000 and 2001. This implies that the increase of space is faster than that of households (which are growing at 1.2%) and leads to a lower price. Figure 4.5 portrays the magnitude and spatial pattern of space price for a typical single family detached house in 2000 and 2020 in the Bay Area. The spatial distribution of price in 2020 is similar to that in 2000. In other words, the decline of the price is proportional in the majority of LUZs. The model outputs are the same as expected.

5. Demonstration Policy Test (SG)

5.1. Purpose

This scenario is designed to test if the model can converge in each forecast year at a higher travel cost. The price is about five times the gasoline price in 2000 and is applied to all years. We are testing that the model can deal with such a high gasoline price and show substantial difference in activity allocation. Our main purpose is to test the performance of the model with an extreme input, so that when realistic scenarios are developed for the production model we can be confident that the model will be able to handle them. In other words, the inputs to this scenario are not intended to be realistic nor do they represent any endorsed policy.

5.2. Development

All the inputs of this scenario are the same as the Base Case scenario except the travel cost. The high gasoline price largely increases the unit travel cost based on distance and has little impact on the unit travel cost based on time. We recalculated the distance-based unit travel cost. Thus, the high gasoline price will affect all choices at the three levels.

5.3. Scenario Results

The figures below show selected results from this scenario, sometimes comparing them with the base case scenario. These figures are discussed below.

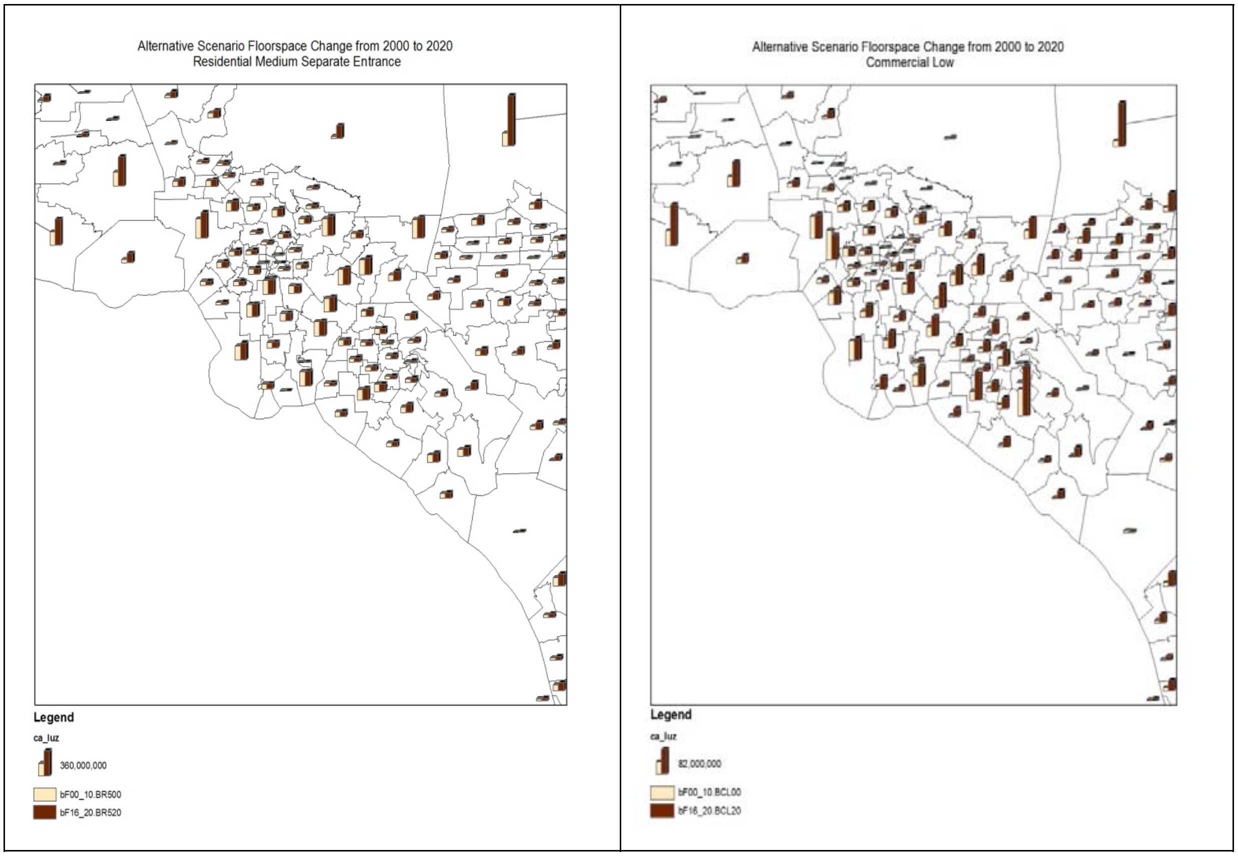

Figure 5.1 Residential medium separate entrance floor space (left figure) and commercial low floor space (right figure) in 2000 (left bar) and 2020 (right bar), Gasoline Scenario, Los Angeles area

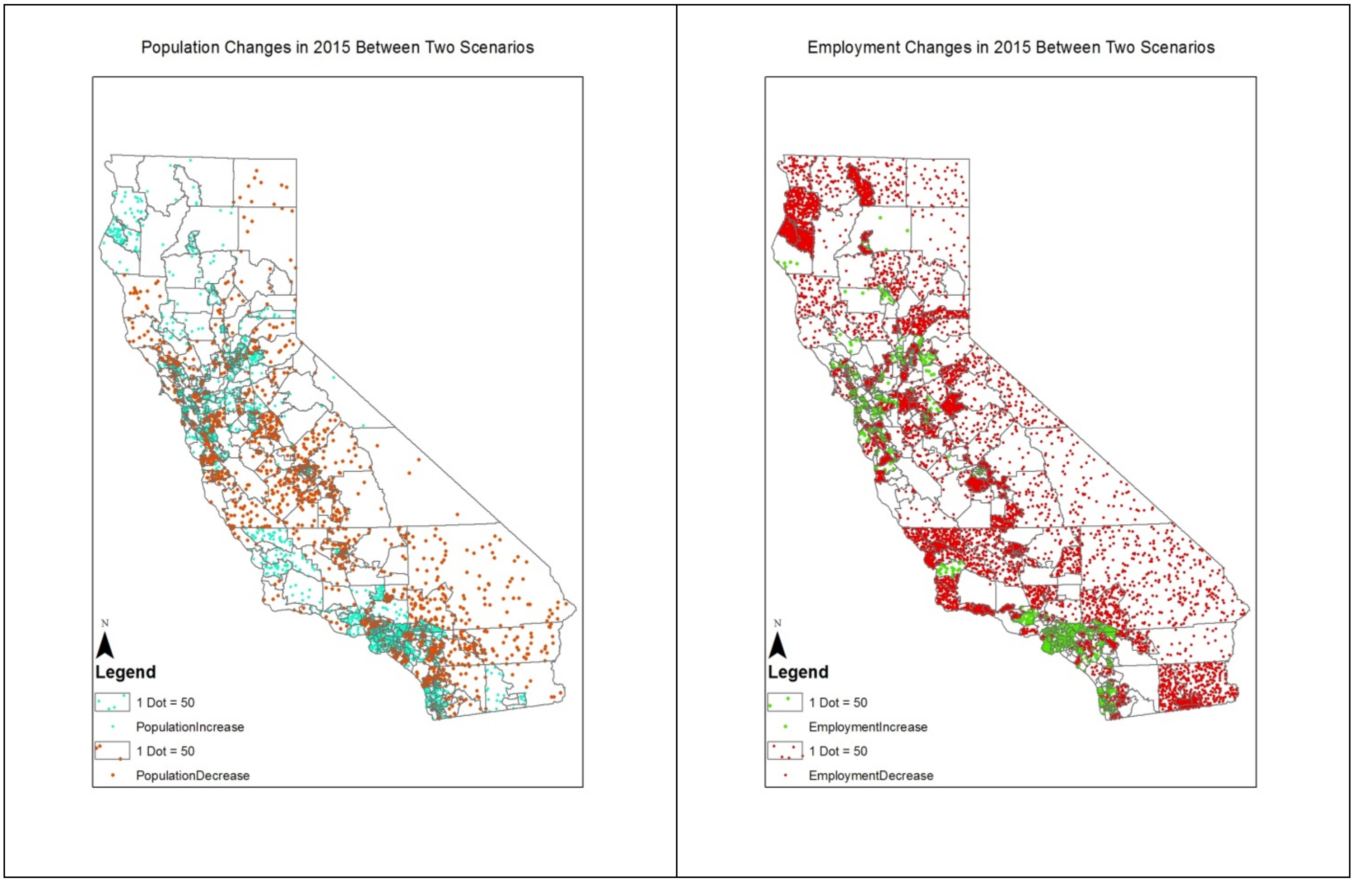

Figure 5.2 Population and employment changes in 2015 between the Base Case and Gasoline scenarios

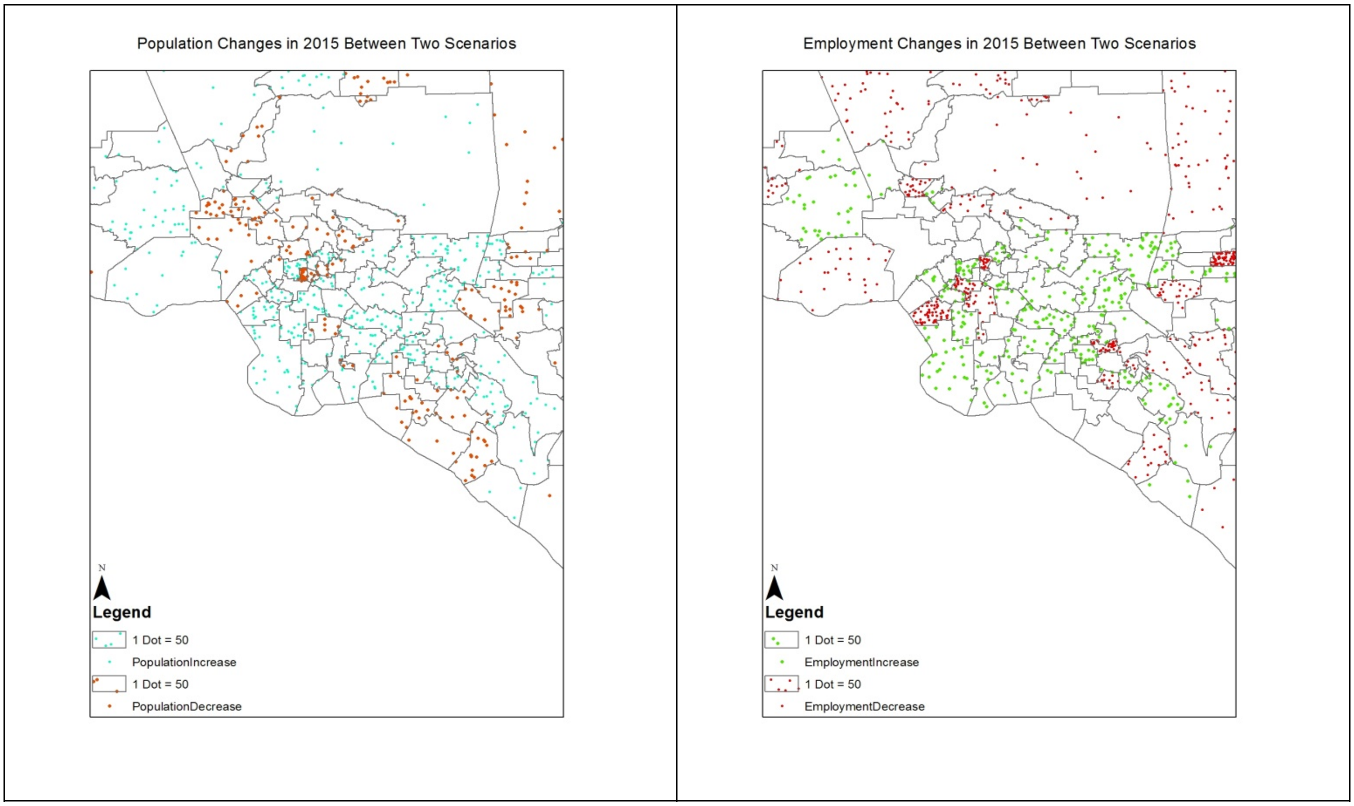

Figure 5.3 Population and employment changes in 2015 between the Base Case and Gasoline scenarios, Los Angeles

Our expectations regarding this scenario are as follows:

- The higher travel cost would make development more compact and closer together, enabling the same interactions with shorter trips;
- The average trip lengths for services (business) trips and work trips (commuting) would be shorter; and
- Floor space prices in zones with higher accessibility would be higher.

From Figure 5.1, we can see that the general land use patterns in the Gasoline scenario are similar to those in the Base Case scenario, i.e., residential (Residential Medium Density Separate Entrance) has a higher growth rate in the suburbs while Commercial Low Density has a higher growth rate in the inner urban areas.

Figure 5.2 shows the absolute differences between the two scenarios in households and employment for the entire state. The dots in blue and green indicate an increase in households and employment, respectively, in those zones. The dots in brown and red represent decreases. In general, the model predicts the hypothesized changes. Certain urban areas grew more quickly because trip lengths can be shorter in urban areas. But the situation is fairly complex, with employment sometimes following population differences and other times in the opposite direction of population changes. This is due to the richness of the model in its representation of different industries. For some industries, it is very important for them to be near their industrial suppliers and industrial customers, and an increase in transportation costs will drive these industries closer together so that they can still take advantage of agglomeration economies. Other industries are tightly tied to population, either for their labor force or for their customers, and higher transport costs can drive them outwards to be closer to the population that they employ and serve. In general, employment seems more sensitive to the higher travel cost than households, and the effects are different between different industries.

Figure 5.3 shows more details in the differences between the two scenarios in Los Angeles area. Comparing figure 5.3 with figures 4.4 and 5.1, we can see that, albeit households grow faster in the suburbs than in the inner urban areas in the Gasoline scenario, they grow more slowly than in the suburbs in the Trend scenario. Further, Los Angeles has a higher growth rate overall in the Gasoline scenario, capitalizing on its central role in the economy of the state. This implies that the model forecasts the growth as expected.

6. Conclusions

The model system has been calibrated to provide a good fit to targets as intended. Its behavior in response to the policy test considered is both understandable and reasonable.

The project has provided a clear indication that a PECAS model of California can be developed to provide a practical tool for policy analysis covering a wide range of alternatives. In fact, the system and approach show great promise. It is recommended that the work continue into the development of a production version of the model as has been proposed.