WESTERN CAPE LAND USE MODEL OVERVIEW

P DONOSO[1], L GUTIÉRREZ[1], F SAAVEDRA[1] and F CONTIERO[2]

[1]Universidad de Chile, LABTUS, Blanco Encalada 2120, Piso 5, Santiago, Chile
Tel: 0056 22 9784380; Email: pedrodonosos@gmail.com
[2]Bentley Systems, Inc., Carl-Zeiss-Ring 5, 85737 Ismaning, Germany
Tel: 0049 1728837391; Email: filippo.contiero@bentley.com

ABSTRACT

The Western Cape Land Use Model (WCLM) is integrated with the Western Cape Transport Model and Western Cape Freight Model in a system constituting the Western Cape Land Use and Transport Interaction Model (WCLUTI).

The WCLM includes several model components: regions are stratified by inside and outside the City of Cape Town, markets are divided by residential and non-residential, and land use identity is divided by “formal” and “informal”.

The “formal” component is modelled using *CUBE Land* software, which is designed to forecast the expected occupied real estate supply, location of households and firms/jobs, and rent values. The “informal” residential land use (additional informal dwellings and informal settlements) is modelled by means of utility-based supplemental models, affecting the available land and the formal location behaviour in areas close to informal property groups through a feedback loop.

The objective of the joint model is to predict these components for the forecasting year under different user-defined scenarios of population and economic growth, regulations and subsidies, real estate projects and accessibility/attractiveness levels provided by the transportation system.

These forecasts feed the transportation model in an integrated system, informing the land use and transport planning process through performance-based indicators, helping to achieve coordination across different sectors.

1. INTRODUCTION

The Western Cape Land Use Model (WCLM) is a land use model of the Western Cape Province, that provides a digital representation of the spatial allocation of total households and total employment in the province, stratified by modelled category. The principal function of the WCLM is forecasting these control totals by agent category, real estate type, and zone in the study area, under different future scenarios of population and economic growth, regulations, subsidies, real estate projects and transportation projects.

The WCLM is not a demographic model, i.e., it does not estimate births, deaths, immigration, or emigration of agents by year, hence it is not modelling provincial total growth. The WCLM is instead allocating this growth spatially, at zonal and real estate type level.

The WCLUTI, of which WCLM is a part, aims to become a quantitative planning instrument for the Western Cape Government for testing and analysing the impact of transport and land use policies and projects. The WCLM was designed to represent interaction of the land use system with the transport system. The WCLUTI model is still under development at the time of writing this paper and has been documented in other papers presented at the SATC (Allen & Martheze, 2021).

2. WCLM GENERAL STRUCTURE

The WCLM aims to characterize the complex land use system in the province, in a flexible structure that allows to accommodate model refinements and overcome data availability constraints.

2.1 Study Area and Regions in WCLM

The Western Cape Province (WC) is the study area and covers 129,462 square kilometres formed by six districts: City of Cape Town (CCT), West Coast, Cape Winelands, Overberg, Garden Route and Central Karoo. Data from the 2016 Community Survey shows that 64% of the inhabitants are concentrated in CCT, but this district only represents 2% of the land territory and the remaining 98% contains large rural areas. Due to this great heterogeneity, two regions have been modelled separately: inside and outside the City of Cape Town.

In the first model version, control totals of households and employment by region are scenario-specific inputs and no interaction between developers and consumers of both regions is assumed. Future refinements of the model might allow for the simulation of an "in-out / out-in" migration mechanism between the two regions (and even these areas could be subdivided).

2.2 Sub-Markets in WCLM

Two sub-markets are considered in each region: residential (modelling allocation of households) and non-residential (modelling allocation of employment). Each location is assumed to be occupied by only one household or employee. Mixed-use properties, occupied by more than one agent type, are subdivided or assigned to the predominant use.

2.3 Inclusion of the informal Market in WCLM

Most types of land use software, including *CUBE Land*, simulate the operation of the formal real estate market with established rules: properties are legally built by developers who seek to maximise their profits; consumers choose location options that maximise utilities and pay a purchase or lease price (rent) for them. But in part of WC, the Land Use system operates informally, where these rules do not apply or only apply partially.

Multiple types of informal dwellings can be identified in WC, with transition occurring between categories and a high degree of dynamism in the system. Nevertheless, two types of informalities were identified for a simplified representation of the system in the WCLM:

- **Additional informal dwelling** (sometimes called "backyard dwelling"): a secondary dwelling in a formal housing unit which is occupied by a household other than the household inhabiting the main dwelling. The main properties are legal units, whilst the

secondary unit is illegal and produced by the main household. We assume that the household that occupies the secondary property pays a rent for it.

- **Informal dwelling in a settlement**: a substandard housing unit, typically illegally located on land belonging to an institution or person other than the person living in this property. These are independent and generally precarious units, and it is assumed that no rent is paid for them. In certain cases, the Anti-Land Invasion Unit is entitled to demolish the informal properties. Otherwise, temporary services might be provided, and these units might be able to upgrade to formal units in the long term through housing programs.

As these properties cannot be modelled within *CUBE Land*, supplemental models were developed and integrated in the WCLM.

2.4 General Structure of WCLM

The WCLM is structured by region (inside and outside CCT), by submarket (residential and non-residential) and by land use identity (formal unit, additional informal dwelling and informal dwelling in a settlement). The model includes three major sub-models: real estate production models (supply), household/employee location models (demand) and real estate price models (rent).

Figure 1 shows the structure of the demand and supply models associated with the above disaggregation groups. For simplicity, only the sequence of calculations (inputs/outputs) is represented in the diagram but not the interactions between the components.

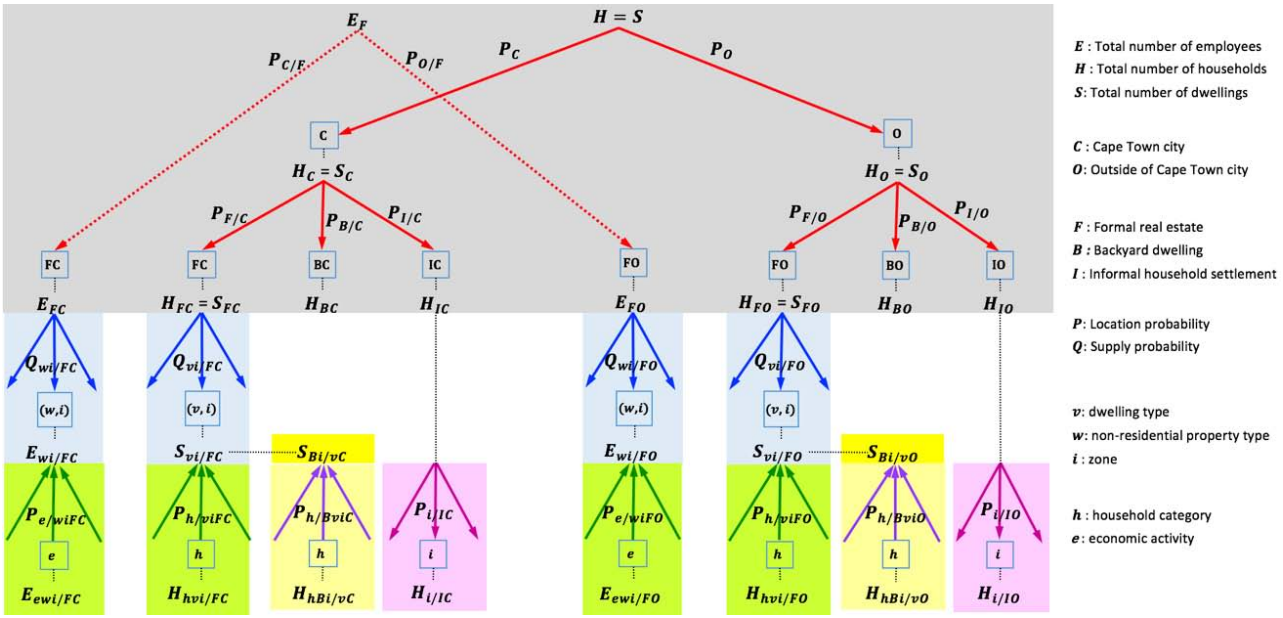

Figure 1: Structure of the WCLM

The totals of households and employees in the province in the forecast year are the overall starting points of the WCLM simulation and a user input for the specific scenario (these totals, disaggregated by household category or economic activity, are needed in subsequent models as well). The WCLM spatially distributes these grand totals throughout the province by zone, real estate type and agent category. The WCLM simulation consists of the following three stages, where the results obtained in the last two stages depend on the results of the corresponding first stage:

1. Disaggregation of the total number of households, with factors calculated from available observed data, into 3 groups: those located in formal properties (grouped by type *v* and zone *i*), those located in informal additional informal dwellings and those located in informal settlements by zone (grey box in Figure 1).
2. Generation of location estimates in the formal market by applying supply and location discrete choice models (light-blue and green boxes in Figure 1, respectively) with the *CUBE Land* software.
3. Generation of supply and location estimates in informal additional informal dwellings (yellow and light-yellow boxes in Figure 1, respectively) and location in informal settlements (Light-purple box in Figure 1) by applying supplemental discrete choice models.

Supply models allow to capture the factors affecting the developers' profit from developing housing or workplace units in the formal market, and the profit for a formal dwelling owner who builds an additional informal dwelling inside his/her property. Demand models allow to capture the factors that affect the utility of the household or the firm by locating its members or its employees, respectively.

The rent/price models in the formal market and for additional informal dwellings are not shown in Figure 1, but they are implicitly considered, because the formal supply models (light-blue boxes in Figure 1 and the additional informal dwelling supply models (yellow boxes in Figure 1) depend upon developer's or owner's profits, which are functions of rents/prices.

2.4.1 Formal Land Use Model with CUBE Land

The land use modelling of the formal market (second stage of the WCLM simulation) is done with *CUBE Land*. This software package was originally developed by the University of Chile and applied in Santiago of Chile under the name of MUSSA (Martínez and Donoso, 2010). CUBE Land is distributed by Bentley Systems and integrated in the *CUBE* software platform, providing GIS, data processing and travel demand modelling modules. *CUBE Land* is a computational representation of a Land Use system and its real estate market. The software simulates the behaviour of real estate developers and households/firms as consumers, in the residential and non-residential markets.

A *CUBE Land* model forecasts occupied real estate supply and rents for different types of properties and zones and the location of different categories of households and types of firms/jobs in these supply options throughout the study area. These forecasts are obtained assuming a static market equilibrium condition of the Land Use system (all consumers are located in the available properties in the zones) and under a variety of scenario conditions of population/economic growth, regulations, subsidies/taxes, real estate projects and access levels provided by the transport system.

The interaction between transport and land use systems allows the applier of the model to assess the impact of projects and policies on both systems. On one hand, the real estate market defines the location of households and firms, classified into agent clusters and real estate types and zones, which drive the transport demand. On the other hand, public and private transport systems set zonal access levels, which represent the advantages/ disadvantages of the zone for making or attracting trips. These indicators influence real estate demand and prices, and thus influence the real estate production.

In the field of real estate economics, the allocation process has been modelled by treating consumers in two alternative ways, as *choosers* (price takers) or *auctioneers* (price makers). *CUBE Land* employs the auction approach, considered a suitable method for modelling the location decisions due to the quasi-unique nature of the goods in the real estate market. A property is a good composed of a building and a location and while the first component can be replicated, the second is not.

Under this auction-based approach the consumer's preferences are represented by bids. Formally, the location bid of an agent is the willingness to pay for a property to achieve a certain level of utility. The bid function is obtained by inverting, in the real estate price, the indirect utility function conditional on location (Solow, 1973; Rosen, 1974). Therefore, the willingness to pay can be understood as the monetary value of the utility, and it can be shown that the utility maximisation paradigm is equivalent to the bidding approach under which suppliers maximize their profit by selling/renting to the best bidder (Martínez, 1992). The willingness to pay is the monetary value of consumers' preferences and is represented by a function, specific of each agent, of attributes of the property and its location. For example, the willingness to pay of one household category might be higher than that of other category for locations in neighbourhoods closer to better schools and transport facilities.

Each bid function is modelled in *CUBE Land* as a stochastic function, expressed as the sum of a deterministic component and a stochastic error (unknown part of the bid). The deterministic part depends on different variables like agent characteristics, property attributes and zonal indicators, and the stochastic one is assumed to be identical, independent and Gumbel distributed. With these statistical assumptions, the location probability of an agent is given by a Multinomial Logit (MNL) model. Based on these location probabilities, the total number of agents of type *h* located at properties of type *v* in zone *i* (H_{hvi}), is calculated as follows:

$$H_{hvi} = S_{vi} P_{h/vi}((B_{gvi})_g) \quad \forall h, v, i \tag{1}$$

where:

S_{vi} Number of location units of real estate type *v* in zone *i.*

$P_{h/vi}$ Location probability of agent *h* being located in property *v* in zone *i* (corresponding to the probability that the bid of this agent is higher than the bids of the other agents, i.e. it is a probability of choosing agents).

B_{gvi} Deterministic part of the bid function of agent *g* for real estate type *v* in zone *i*.

Following the auction approach, the rental price (r_{vi}) of a real estate type *v* in zone *i* is determined by the maximum bid among all bidders. As can be seen, bid functions define location and rent values simultaneously, which is a noteworthy property of *CUBE Land*. Since rent values are estimated endogenously in this model (resulting from latent bids), consumers act as price makers.

On the supply side, the maximum profit paradigm is applied, i.e. developers choose what type of real estate to produce (type of property *v* and zone *i*) to maximise profit, restricted to comply with capacity constraints and zoning regulations. The profit is modelled as function of real estate prices/rents and costs and a stochastic term assumed to be identical, independent and Gumbel distributed. Then the production probability of a location option (*v,i*) is a Multinomial Logit model and the total number of type-*v* properties in zone *i* is given by:

$$S_{vi} = HP_{vi}((r_{wz} - c_{wz})_{w,z}) \quad \forall h, v, i \tag{2}$$

where:

S_{vi} Number of location units of real estate type *v* in zone *i* in the forecasting year.
H Total number of agents.
P_{vi} Probability of supplying property *v* in zone *i*.
r_{wz} Rent/price of real estate type *w* in zone *z*.
c_{wz} Supplying cost of real estate type *w* in zone *z*.

Therefore, the core functions of the *CUBE Land* models – driving the real estate production, location and rent formation – are the bid and cost functions. These functions depend on agent, property and zonal attributes and parameters to be estimated; the most common specification is a linear-in-parameters function.

The microeconomic rationale of *CUBE Land* allows for different regions in the study area and sub-markets (*m*) in each region to differentiate consumers' and developers' behaviours. Each sub-market is characterised by its own portfolio of properties and consumers that can only be located in one of the supply options of the portfolio. This strategy recognises that properties are different because they are built for different activities and that there are different regulations that restrict their production and occupancy. Consistently the supply, location and rent models vary in the sub-markets. However, all supply options from all sub-markets compete for the same space in the region, under the existing land use regulations, and this is the first linkage between different sub-markets. In the WCLM two large sub-markets are considered, the residential and the non-residential sub-markets: the first one formed by dwellings and households and the second one by workplaces and employment.

The formulation of *CUBE Land* models follows an aggregate discrete approach for all units of demand and supply: households and firms/employment are clustered into categories (*h*), while land is divided into zones (*i*) and properties into types (*v*). This strategy reduces the dimensionality of the land use equilibrium problem that is solved in a *CUBE Land* simulation. The modeller defines the zoning system, the demand categories and the supply types, with flexibility in terms of categorisation and dimension. Since the observed available information on employment in WC is not classified by real estate type (*v*), this supply dimension is omitted in the specification of the employment model. This condition can be removed in the future, if detailed data is available.

Finally, it is worth mentioning that *CUBE Land* can represent the influence of location externalities in the agents' preferences. These factors correspond to zonal attributes included in the bid functions that depend upon the residential and/or non-residential location. An example of such an externality is the zonal average residential income. Specialised iterative algorithms are implemented in the software for estimating internally the values of the externalities.

2.4.2 Informal Land Use Supplemental Models

The last stage of the WCLM deals with the informal residential land use in the Western Cape. To do this, supplemental models were developed outside the *CUBE Land* software but integrated into the CUBE platform through script-based *CUBE Voyager* programs. In each region (inside and outside CCT), the following supply and location models of additional informal dwellings and a location model of informal settlements were implemented. For the sake of simplicity, region sub-indexes are omitted since the same model structure was used in each region.

The supply model of additional informal dwellings forecasts which formal dwelling is likely to be used by an additional household living in the informal attached property, with the following functional form:

$$S_{Bi/v} = H_B * \frac{S_{vi} Q_{B/vi}}{\sum_{v',i'} S_{v'i'} Q_{B/v'i'}} \quad (3)$$

where:

$S_{Bi/v}$ Number of additional informal dwellings *B* in type-v property and zone *i* in the forecasting year.
H_B Total number of households living in additional informal dwellings *B* in the forecasting year in the region.
S_{vi} Total number of formal type-v dwellings in zone *i* in the forecasting year.
$Q_{B/vi}$ Probability of producing an additional informal dwelling *B* on a type-v formal dwelling in zone *i* in the forecasting year, with respect to the condition of not producing it.

The probability $Q_{B/vi}$ of producing a additional informal dwelling *B* in type-*v* property and zone *i* in the forecasting year is represented by a binomial logit model or a logistic regression given by:

$$Q_{B/vi} = \frac{\exp(\tilde{\pi}_{vi})}{1 + \exp(\tilde{\pi}_{vi})} \quad (4)$$

where $\tilde{\pi}_{vi}$ corresponds to the benefit perceived by the household of the main dwelling from producing a secondary dwelling *v* in zone *i*. This profit depends upon the rental and cost of this unit.

As additional informal dwellings are rented, the location model for these properties has a similar MNL functional form to the one used for the formal location model.

Regarding informal settlements, it is assumed that the location model is a choice that makes a household, of the place to build an informal unit and to be located, therefore the choice of the zone amongst all zones that (could) contain settlements. The location "preferences" will be mostly explained by attributes that assure minimum living conditions, spatial connectivity, or proximity to services and employment areas. Following this approach, the location probability model is assumed to be Multinomial Logit:

$$P_{i/I} = \frac{\exp(\mu' \overline{U}_{i/I})}{\sum_{i'}^{Z} \exp(\mu' \overline{U}_{i'/I})} \quad \forall i \quad (5)$$

where:

$P_{i/I}$ Probability of a household being located in zone *i* containing informal settlements *I* in the forecasting year.
μ' Factor scale of the Gumbel distribution.
$\overline{U}_{i/I}$ Deterministic utility function component of a household for being located in an informal settlement in zone *i* in the region in the forecasting year.
Z Number of zones in the region containing a settlement in the forecasting year.

3. WCLM IMPLEMENTATION

In this chapter, different aspects of implementation of the WCLM in the Western Cape province are analysed.

3.1 Data Availability to Implement WCLM

Data collection for the estimation of the Demand/Location, Rent/Price and Development/ Supply models, was a key challenge in WC. Indeed, the necessary information could not be obtained from a single data source. Table 1 provides the current status in terms of data collection for the estimation of the model. The table contains the main data sources, but other data have been collected and used to derive additional data. It is important to highlight that *supply* data refers to real estate as *goods*, whilst location (*demand*) data refers to the *use/occupancy* of these goods.

Table 1: Summary of the Main Data Sources for the Model Estimation

Spatial level	Market / Land Use	Residential (Households by cluster)			Non-Residential (Employment by economic sector)		
		Locations ("who" and "where")	Rents/Prices (related to locations)	Supply (development)	Locations ("who" and "where")	Rents/Prices (related to locations)	Supply (development)
		Occupancy of RE		RE as *good*	Occupancy of RE		RE as *good*
Inside CT	Formal	Aggregate: Census 2011 + General Valuation Rolls (GVR) 2015	Disaggregate: rounded tax value Rode's survey TPN data	Aggregate: Census 2011	Aggregation: GVR → Employment Braby's data CCT Empl Survey	Disaggregate: rounded tax value Rode's survey TPN data	Aggregation: GVR → Employment
	Informal: Additional + Settlement	Aggregate: Census 2011	-	Aggregate: Census 2011 + "Pocket Areas/Points"	-	-	-
Outside CT	Formal	Aggregate: Census 2011	Disaggregate: purchase price Rode's survey TPN data	Aggregate: Census 2011	Aggregation: Employment Braby's data	Rode's survey TPN data	Aggregation: Employment
	Informal: Additional + Settlement	Aggregate: Census 2011	-	Aggregate: Census 2011 + "Pocket Areas/Points"	-	-	-
Colours Legend:	Estimated/Synthetic data	Procured data	Partial data	No data / no usable data			

3.2 Tuning of the WCLM According to the Available Information

Not all the identified components and disaggregation levels represented in Figure 1 could be included in this first version of the model, due to missing data. The estimated models by stage and the aspects excluded or modified are indicated below.

Stage 1:

- All control totals are estimated using factors obtained from 2011 Census and 2016 estimated employment data.

Stage 2:

- The non-residential models, which are estimated using 2016 employment data, allocates employees only in formal real estate units due to the lack of information on informal non-residential location. Moreover, available employment data cannot be classified by real estate type, thus non-residential supply and location models cannot be disaggregated by this dimension.
- It was not possible to estimate the rent models with the available data, nor the formal supply models. Therefore, the number of formal dwellings by type and zone and the total number of employees by zone are user-defined scenarios, restricted to match totals resulting from Stage 1. This aspect is currently under revision in the model formulation.

- The residential location models of formal dwellings are estimated using 2011 Census data.

Stage 3:

- The additional informal dwelling supply and location models are estimated using 2011 Census data. Unfortunately, it is not possible to link these secondary properties to the main formal properties, but only to the zone in which both are located. Therefore, dimension *v* in equations 3 and 4 must be discarded. This restriction prevents the inclusion of attributes of the main dwelling that could influence the supply and occupancy of additional informal dwellings. Additionally, there is no information on additional informal dwelling rental values that would allow us to improve the supply models of these secondary properties.
- The models for locating households in settlements are estimated using 2011 Census data.

Despite the limitations in terms of data availability, the WCLM represents the main components characterising the complex land use system in the province, but also includes the flexibility to accommodate future model refinements and additional data collection.

3.3 Supply and Demand Clusters in WCLM

The supply and demand clusters define the resolution level of the model and the simulation results. The clusters should be defined based on the variables that allow the user to distinguish location options and agent types associated to different location patterns. Nevertheless, the limited data availability strongly constrained their definition for the WCLM.

3.3.1 Agent Types in the Residential Sub-Market: Household Categories

Regarding the residential sub-market, the Census 2011 represented the only available source to categorise households segmented by housing types, using the Stats SA cross-classification software (*super-cross*). Considering that it was not possible to use "vehicle ownership" because it was not possible to cross-tabulate this variable, the household categories that can be defined in the WCLM are based on the following two attributes: annual household income and household size (number of persons in the household).

Regarding the *income* attribute, the clusters were defined with the purpose of recognising different location patterns within homogeneous categories, with a categorisation *approximately* consistent with the Western Cape Transport Model (WCTM) structure (Allen & Contiero, 2021).

Table 2: Number of households by income category adjusted by income imputation for the WC Province

Income category		Total	%
Low income	**R 1 - R 19 600**	336,971	22%
Medium-low income	**R 19 601 - R 38 200**	321,437	21%
Medium income	**R 38 201 - R 76 400**	284,216	18%
Medium-high income	**R 76 401 - R 153 800**	230,969	15%
High income	**R 153 800 or more**	392,683	25%
Total		**1,566,276**	**100%**

The Census 2011 data shows that most households living in additional informal dwellings and informal settlements belong to the 3 household categories with lowest income level. It has been assumed in the model that households located in informal units belong to only one of the first three income categories.

Regarding the household *size*, groups are defined using the same criteria employed in the definition of income categories and they are listed in the following table.

Table 3: Number of households by size levels for the WC province

Household size (number of members)	Total	%
1	389,959	25%
2	333,611	21%
3	247,659	16%
4	248,956	16%
5	150,341	10%
6 or more	195,795	13%
Total	**1,566,276**	**100%**

Therefore, the WCLM uses 30 household categories in total, resulting from the combination of all income categories (5) and all size categories (6).

3.3.2 Agent Types in the Non-Residential Sub-Market: Employment Categories

In the case of the non-residential sub-market, two primary data sources have been adopted for the segmentation of the employment in economic activities: GeoTerra Image (GTI) 2015 data for the entire WC province, General Valuation Rolls (GVR) – Gross Leasable Area (GLA) 2015 data inside the CCT only. To derive suitable data though was a complex process involving the processing and integration of multiple data sources. Seven large employment segments were established according to economic activity and the available data, avoiding small groups:

- Agriculture + Mining (SIC_1 + SIC_2): Agriculture, forestry, fishing + Mining, quarrying.
- Manufacturing (SIC_3): Manufacturing.
- Construction (SIC_5): Construction.
- Trade (SIC_6): Wholesale and retail trade, catering and accommodation.
- Transport (SIC_7): Transport, storage and communication.
- Finance + Utility (SIC_8 + SIC_4): Finance, insurance, real estate and business services + Electricity, gas and water.
- Government + Social (SIC_9): General government + Community, social, personal services.

Some of these segments are formed by combining two different SIC codes. This was done to avoid forming segments that are too small and to cluster activities that were showing a similar localisation pattern. This occurs in the case of *Finance* and *Utility* because of the size of the groups and similar location patterns observed from the data. Further analysis is currently undergoing for these categories, with more detailed supplemental treatments of some of the more constrained components.

3.3.3 Real Estate Types in the Residential Sub-Market

The Census data provides two dwelling attributes that are used to define housing types: main dwelling type and number of rooms.

The housing types based on these two variables were defined considering different criteria: separating formal and informal dwellings, distinguishing different production and occupancy patterns of formal real estates and avoiding very small and different-sized types when possible. Since the *Other* type of housing cannot be treated as a type to be modelled, it is distributed among the other five types in proportion to their shares. By doing

this, the following frequency table is obtained (note: these are not the final categories used in the model, as explained below).

Table 4: Number of dwellings by modelled housing types for the Western Cape Province

Grouped main dwelling type		Number of rooms			Total	%
		1-2-3	4-5	6 or more		
Formal	Single	253.325	388.618	344.382	**986.325**	**63%**
	Multi	60.134	85.986	16.240	**162.361**	**10%**
	Flat	44.895	72.614	6.998	**124.507**	**8%**
Additional informal dwelling		92.492	10.059	505	**103.055**	**7%**
Informal dwelling		153.366	35.454	1.208	**190.027**	**12%**
Total		**604.212**	**592.731**	**369.333**	**1.566.276**	**100%**
		39%	**38%**	**24%**	**100%**	

Based on this table, 11 housing types were defined as a result of dividing the dwelling type *single* into the 3 number-of-rooms-based types (3 or less rooms, 4 or 5 rooms, and 6 or more rooms), and the remaining dwelling types (multi, flat, informal additional informal dwelling and informal dwelling - mostly settlements) only into 2 number-of-rooms-based types (3 or less rooms, 4 or more rooms). Then, some cells of this table were grouped to avoid defining very small housing types.

3.3.4 Real Estate Types in the Non-Residential Sub-Market

As indicated before, the available employment information in WC cannot be classified by real estate type, so it is not possible to include this dimension in the model.

3.3.5 Zoning System

Considering the available data for estimating the land use models, the reference zoning system needed to be based on the Stats SA Small Area Layer (one SAL = one zone).

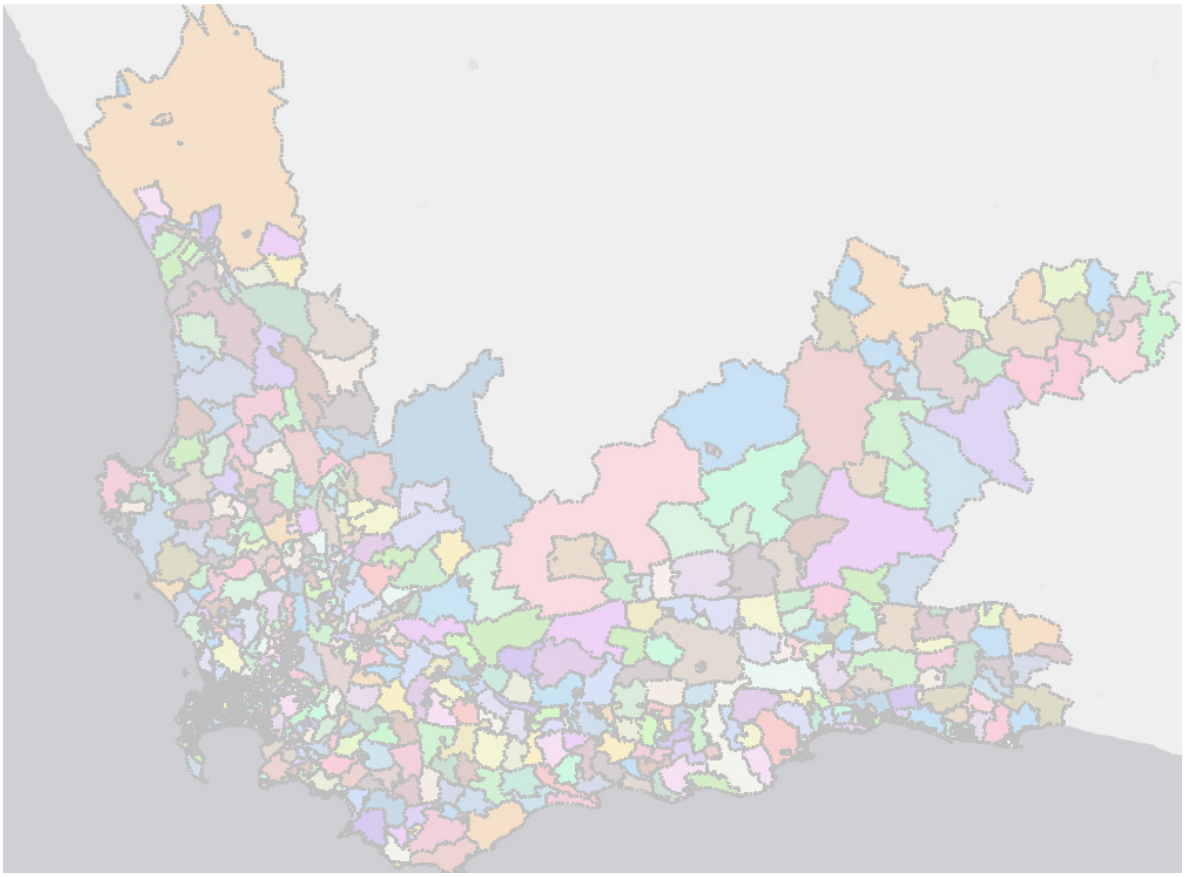

Figure 2: Zone system used in the WCLM (based on Stats SA SAL)

The zoning system for the *CUBE Land* model should be compatible with the WCTM zoning system. A data aggregation-disaggregation procedure has been implemented accordingly, to guarantee consistency between data processed at the SAL level (8,798 SAL zones – modified version of the original SAL zoning system, filling the gaps with additional zones) and at WCTM TAZ level (3,954 TAZ zones).

3.4 Data Preparation for the WCLM Estimation

Although *CUBE Land* works with aggregate demand and supply groups (i.e. household/ employment categories, real estate types, zones), the microeconomic basis of the model makes it possible to specify disaggregate location, rent and supply models (i.e. models applied to individual agents and locations). Disaggregate data usually comes from spatial samples, from which *CUBE Land* models "learn". Since the variability of disaggregate data is usually larger, more parameters and attributes can be identified in individual models than in group-based models with higher statistical significance. Nevertheless, disaggregate data are sometimes difficult to gather, and aggregate data can be used to estimate the components of the model when they are not available, as is the case for the WCLM.

The data to estimate the sub-models can be classified into three main groups: agent attributes (*h,v,i*), property attributes (*v,i*) and zonal attributes (*i*):

- Since location model parameters are differentiated by agent groups, the agent attributes included as variables for the model formulation should correspond to complementary variables not included in the definition of these groups.
- Regarding real estate unit characteristics, attributes that characterise format, structure, size and *quality* of the properties need to be gathered.
- The third important group of attributes explaining location, rent and supply patterns are zonal characteristics describing the surrounding area of a property, including public infrastructure and environmental attributes.
- Finally, the urban and intercity transport system plays an important role in land use, because households/firms decide where to locate considering the accessibility/ attractiveness of people and cargo. The definition of these access measures and testing their significance in the model is a key aspect for the Land Use – Transport Interaction (LUTI). Several measures have been calculated for capturing this interaction in the model.

All the WCLM sub-models to be estimated are in the form of aggregated Multinomial Logit models, with the exception of the additional informal dwellings supply sub-model which is a Binomial Logit model. The Maximum Likelihood method has been adopted for their estimation.

The supply model specification and the estimation of its parameters depends on rental values. Since information on residential and non-residential rents was not available, an attempt was made to use approximate values of rents. Unfortunately, no appropriate models were obtained, showing the importance of rent data to estimate the production of formal real estates in WC. For this reason, at the current state of development, the WCLM can only be applied with exogenous estimates of the supply. Therefore, a dataset containing the formal supply by zone constitutes a scenario required "fixed-supply" input. Then, the model allocates the agents to this pre-defined supply, which can include new developments at zonal level. The usage of the model with "variable-supply", i.e. capability of simulating the actual development with the model, will be included as a future refinement.

A summary of the explanatory variables considered in the estimation of the location models in formal and informal units and of the supply models for additional informal housing is given in the following tables. Future refinement of the WCLM will focus on incorporating more attributes.

Table 5: Attributes of the location models in formal real estate

Attribute type	Attribute value	Attribute definition
Dwelling attribute	1/0 (6-or-more room Single)	1 if dwelling is Single-family housing with 6 or more rooms; 0 otherwise
	1/0 (3-or-less room Single)	1 if dwelling is Single-family housing with 1 or 2 or 3 rooms; 0 otherwise
	1/0 (Multi)	1 if dwelling is Multi-family housing; 0 otherwise
	1/0 (Flat)	1 if dwelling is a Flat; 0 otherwise
Zonal attribute: residential supply externality of formal dwellings	Spatial density of formal dwellings in the zone	Total number of formal dwellings in the zone / Total area of the zone in square meters
Zonal attribute: residential location externality of formal dwellings	Share of low-income households in formal units in the zone	Number of *low-income* households in formal units in the zone / Number of households in formal units in the zone
	Share of medium-low-income households in formal units in the zone	Number of *medium-low-income* households in formal units in the zone / Number of households in formal units in the zone
	Share of medium-income households in formal units in the zone	Number of *medium-income* households in formal units in the zone / Number of households in formal units in the zone
	Share of high-income households in formal units in the zone	Number of *high-income* households in formal units in the zone / Number of households in formal units in the zone
Zonal attribute: non-residential supply externality	Spatial density of employment in the zone	Total employment in the zone / Total area of the zone in square meters
Zonal attribute: non-residential location externality	Trade employment in the zone	Trade employment in the zone
	Spatial density of trade employment in the zone	Trade employment in the zone / Total area of the zone in square meters
	Share of Gov+Social employment in the zone	Government and Social activities employment in the zone / Total employment in the zone
	Share of Trade employment in the zone	Trade employment in the zone / Total employment in the zone
	Share of Fin+Util employment in the zone	Finance + Utilities employment in the zone / Total employment in the zone
	Share of Agr+Mining employment in the zone	Agriculture and Mining employment in the zone / Total employment in the zone
Zonal attribute: transport accessibility	Average AM peak travel time by car from the zone to a destination in WC	Weighted average travel time in morning peak period in private transport from the zone to all zones in Western Cape Province
	Average AM peak travel time by transit from the zone to a destination in WC	Weighted average travel time in morning peak period in public transport from the zone to all zones in Western Cape Province
Zonal attribute: transport attractiveness	Spatial density of AM peak motorized trips to the zone from origins in WC	Total number of private and public trips in morning peak from all zones in Western Cape Province to zone / Total area of the zone in square meters
	Average AM peak travel time by car from an origin within the same municipality to the zone	Weighted average travel time in morning peak in private transport from all zones within the same municipality to the zone

Table 6: Attributes of the supply and location models in informal dwellings

Attribute type	Attribute value	Attribute definition
Zonal attribute: residential supply externality of informal dwellings	Spatial density of additional informal dwellings in the zone	Total number of additional informal dwellings in the zone / Total area of the zone in square meters
	Number of additional informal dwellings in the zone	Total number of additional informal dwellings in the zone
	Natural logarithm of spatial density of households living in settlements in the zone	Natural logarithm of (Total number of households living in settlements in the zone / Total area of the zone in square meters)
Zonal attribute: residential supply externality of formal dwellings	Spatial density of formal dwellings in the zone	Total number of formal dwellings in the zone / Total area of the zone in square meters
Zonal attribute: residential location externality of formal dwellings	Average income of households in formal dwellings in the zone	Average income of households in formal dwellings in the zone
	Share of low-income households in formal units in the zone	Number of *low-income* households in formal units in the zone / Number of households in formal units in the zone
	Share of medium-low-income households in formal units in the zone	Number of *medium-low-income* households in formal units in the zone / Number of households in formal units in the zone
	Share of households of medium-high and high income in formal units in the zone	Number of households of *medium-high and high income* in formal units in the zone / Number of households in formal units in the zone
	Spatial density of households of medium-high and high income in formal units in the zone	Number of households of *medium-high and high income* in formal units in the zone / Total area of the zone in square meters
	Spatial density of high-income households in formal units in the zone	Total number of high-income households in formal units in the zone / Total area of the zone in square meters
Zonal attribute: non-residential location externality	Share of Agr+Mining employment in the zone	Agriculture and Mining employment in the zone / Total employment in the zone
	Share of Trade employment in the zone	Trade employment in the zone / Total employment in the zone
	Spatial density of Construction employment in the zone	Construction employment in the zone / Total area of the zone in square meters
	Share of Manufacturing employment in the zone	Manufacturing employment in the zone / Total employment in the zone
Zonal attribute: transport accessibility	Average AM peak travel time by transit from the zone to a destination in WC	Weighted average travel time in public transport from the zone to all zones in Western Cape Province
	Average AM peak travel time by transit from the zone to destinations in the same municipality	Weighted average travel time in public transport from the zone to all zones within the same municipality
	Average transit fare of trips from the zone to a destination in WC	Weighted average transit fare of trips from the zone to all zones in Western Cape Province
Zonal attribute: transport attractiveness	Number of AM peak motorized trips to the zone from origins in WC	Total number of private and public trips from all zones in Western Cape Province to zone

Some of the adopted explanatory variables are location externalities (zonal attributes depending on location probabilities). In other words, the location decisions of the different agent types are influenced by zonal attributes determined by residential and non-residential location decisions. This makes it possible to represent spatial agglomerations, for instance agglomeration of households by income level or of different economic activities complementing each other. *CUBE Land* simulation provides location estimates consistent with location externalities defined in the specification of the probabilities, by using an internal iterative algorithm.

Once the *CUBE Land* simulation process is executed, the zonal location externalities of the formal market can be computed and then, the residential location in additional informal dwellings and informal settlements, which depend upon the formal real estate market, can be estimated. This mechanism allows the modeller to represent the influence of the formal

market on the informal system (results show that this is statistically significant). The influence of the informal land use over the formal market was not included in this first version of WCLM to avoid complications introducing the feedback loop. This will be part of future refinement of the model.

3.5 Integration of the WCLM in the WCLUTI Model

The WCLM has been integrated with the Western Cape Transport Model (WCTM) and the Western Cape Freight Model (WCFM) within the *CUBE* platform, constituting the Western Cape Land Use and Transport Interaction (WCLUTI) model. WCLUTI is a "what-if" scenario analysis tool, providing the picture of what would happen in the land use and transport systems when changes are introduced to the model inputs. The *CUBE* interface provides the user with a GUI to select or edit input options/parameters and files that are scenario specific for the WCLUTI Model. The user can run just those model components that are necessary to simulate the specific scenario.

4. CONCLUSIONS

The main purpose of the WCLUTI model is to address planning needs, in terms of creating a quantitative foundation for policies and planning projects to support planning and design decisions in the Western Cape Government, together with achieve coordination across departments.

At the time of writing this paper, the model is still under development, and a few key major challenges connected to the development of the WCLM could be summarized with the following bullet points:

- A first key challenge was related to the zoning system definition itself. Indeed, the Stats SA SAL was used due to data availability. The size and boundaries of these zones often do not match the requirements of land use modelling and analysis that could be done with the implemented model, probably because they were defined with a different purpose. For example, some zones are not homogeneous in terms of the types of land use they contain, include areas not usable for the location of homes and firms or do not separate areas that may contain different projects or land use and transportation policies.

- A major challenge was related to the collection of good quality employment data by SIC code at zone level consistent with the Stats SA SAL (or more refined level). It is very challenging indeed to determine a linkage between land use and employment data at this level. Additionally, obtaining representative information on sale or rental prices of market transactions is a pending task that would allow to forecast this important component of the model and also to improve the real estate supply modelling that depends on prices.

- A third major challenge was linked to the data availability and the capability to model the generation of informal residential units and households' preferences for these units. There are some key data that are not available at the time of writing:
  - The housing and occupant information on a dwelling containing and additional informal unit. This information could greatly improve the modelling of the generation and location patterns of these secondary units.

- Settlement characteristics, such as availability of space and basic services and barriers or incentives for occupants to relocate to other camps or formal housing. These attributes that can help to improve the settlement demand model. Additionally, the generation of informal settlements is not modelled but given as scenario setting. More information on the drivers of settlement generation and permanence can facilitate this task.

- A final key challenge was related to the capability of the location patters to show connection with calculated access measures. Transport affects land use, but available information from both sectors shows that this influence is weak, which should be further analysed.

5. REFERENCES

Allen, WG & Contiero, F, 2021. *An Innovative Travel Model System for Western Cape*. 40th South Africa Transport Conference Paper.

Allen, WG & Martheze, D, 2021. *The need for a Provincial Land Use and Travel Forecasting Model*. 40th South Africa Transport Conference Paper.

Anas, A, 1982. *Residential Location Markets and Urban Transportation*. Academic Press, London.

Contiero, F & Allen, WG, 2021. *The Challenge of Socio-Economic Data for the Western Cape Travel and Land Use Model*. 40th South Africa Transport Conference Paper.

Ellickson, B, 1981 *An alternative test of the hedonic theory of housing markets*. *Journal of Urban Economics*, 9(1):56-79.

Martínez, F, 1992. *The bid-choice land use model: an integrated economic framework*. *Environment and Planning A* 24, 871-885.

Martínez, F & Donoso P, 2010. *The MUSSA II Land Use Auction Equilibrium Model*. In Residential Location Choice, Models and Applications. Editors: J. Preston, D. Simmonds, YF Pagliara, Chapter 4, Advances in Spatial Science, Springer, 99-113.

McFadden, D, 1978. Modelling the choice of Residential location. In: Karlquist, A et al. (Eds.), *Spatial Interaction Theory and Residential Location*. North Holland, Amsterdam, pp. 75-96.

Rosen, S, 1974. *Hedonic prices and implicit markets: product differentiation in pure competition*. *Journal of Political Economy*, 62:34-55.

Solow, RM, 1973. *On equilibrium models of urban location*. In: Parkin, M. (Ed.), Essays in Modern Economics. Longman, London, pp. 2-16.