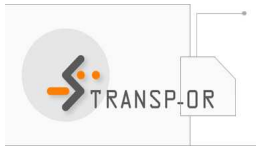
Starting With UrbanSim: On the Creation of an Introductory Project

Olivier Gallay

Ecole Polytechnique Fédérale de Lausanne (EPFL)
TRANSP-OR, Transport and Mobility Laboratory

Internal Seminar
September 23rd 2010



UrbanSim Basics

- Open source urban simulation system (OPUS)
- Support for planning and analysis of urban development
- Interactions between land use, transportation, the economy, the environment, etc
- Python based modularized architecture to facilitate the insertion of plug-ings
- Microsimulation, agent-based, approach

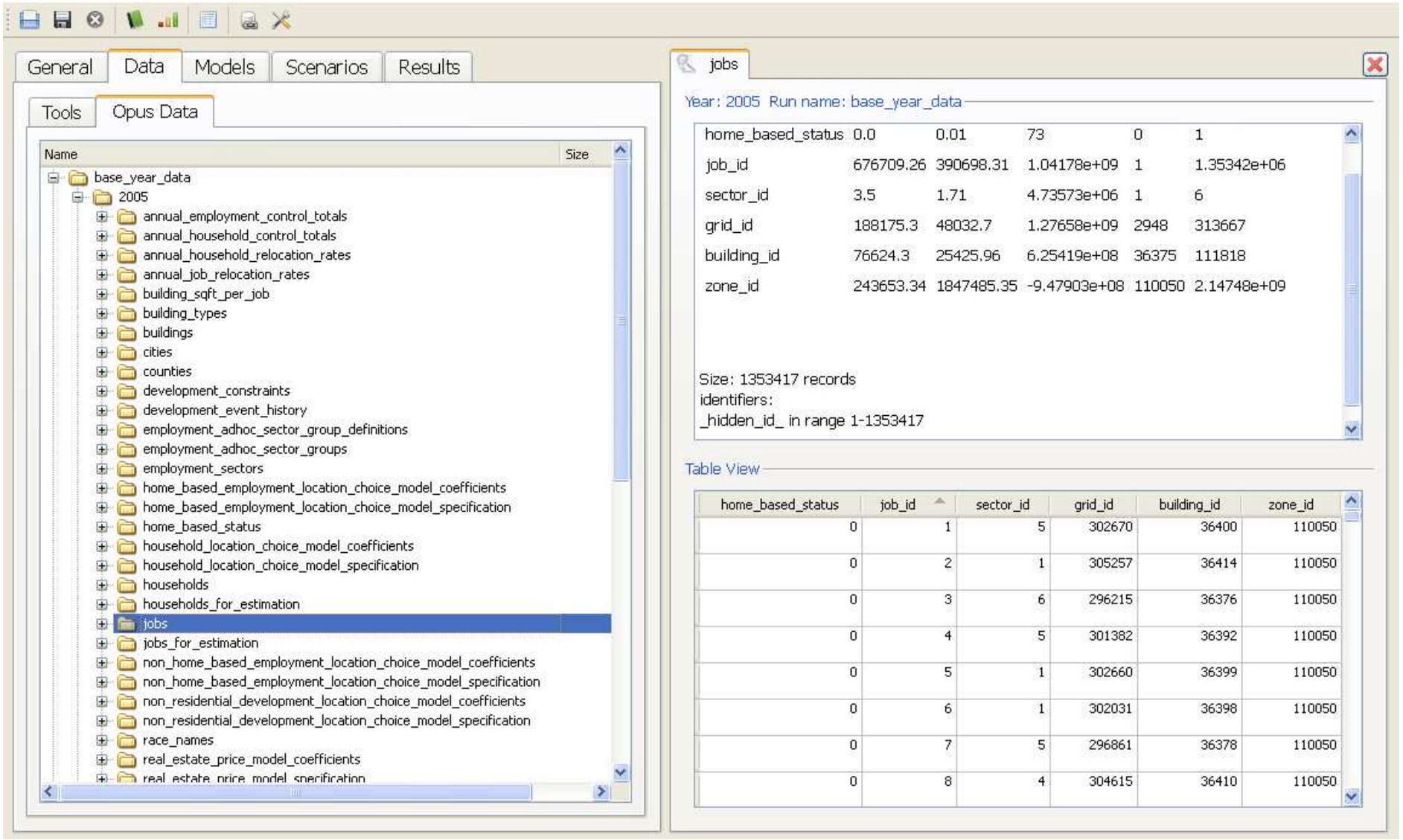
Context

- SustainCity: Brussels case study
- Current dataset not yet available
- Proceed to a preliminary study and identify potential problems arising with the software
- Incomplete (and old) dataset for Brussels
- Use the latest Developper version
- Start from the `san_antonio_zone` project and adapt it to the Brussels dataset

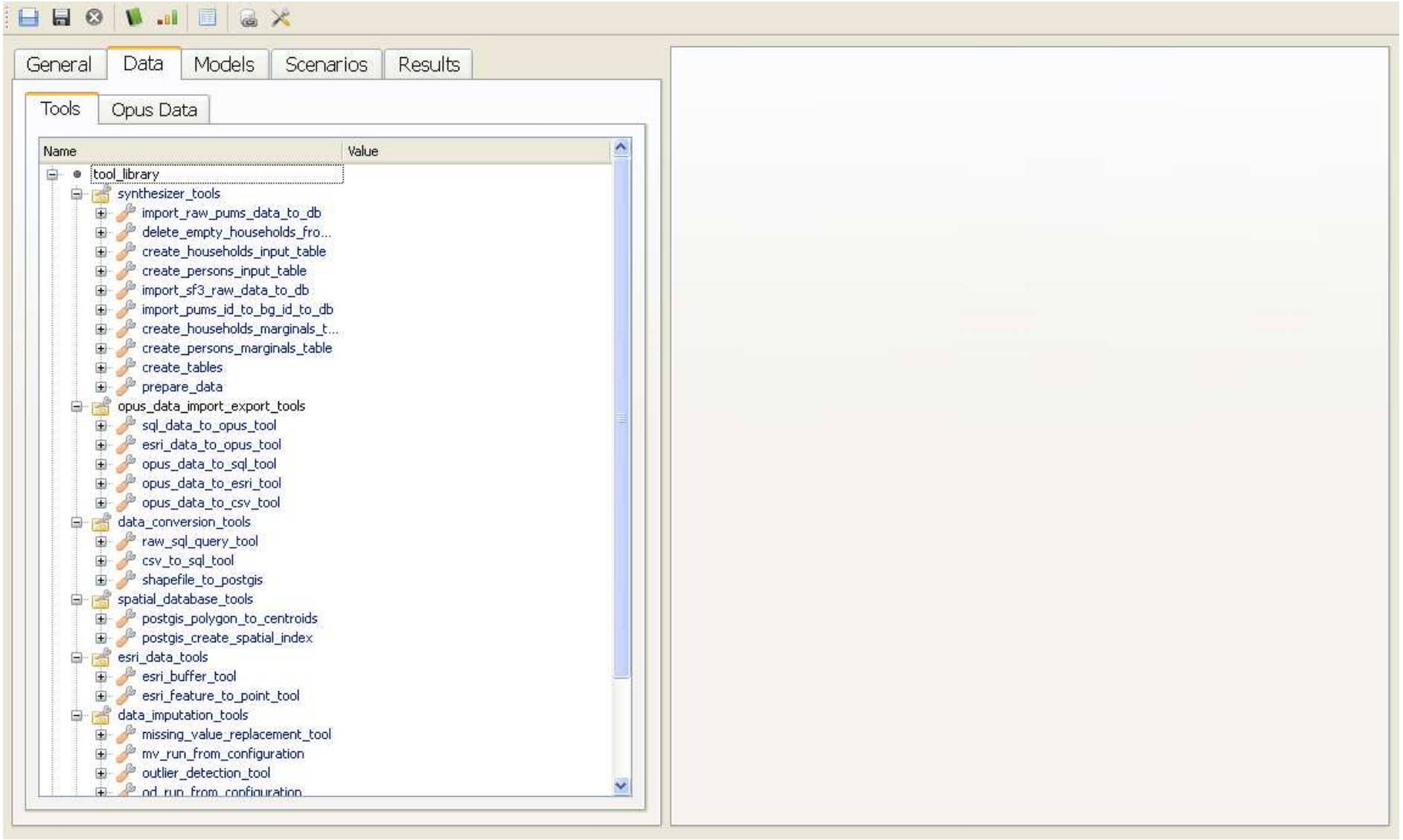
Software Architecture

- Layer construction:
  - `urbansim` (general layer)
  - `urbansim_zone` (zone layer)
  - `brussels_zone` (project layer)
- Strict predominance between layers: child, parent files
- Python code and `.xml` above layer:
  - `urbansim.xml`: model general specifications, database connections,...
  - `urbansim_zone.xml`: model parameters
  - `brussels_zone.xml`: model specific configuration (variable set)

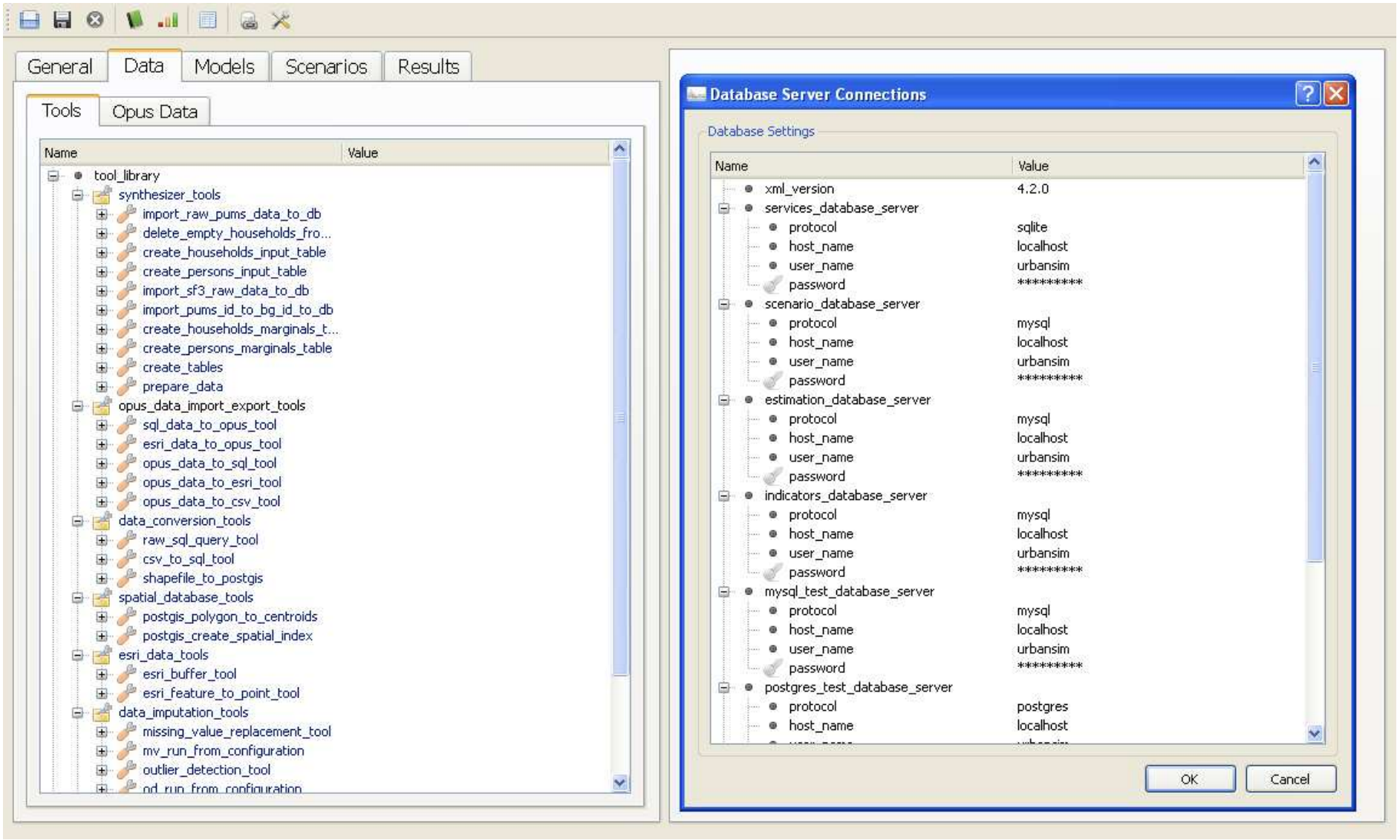
UrbanSim Data Cache Format



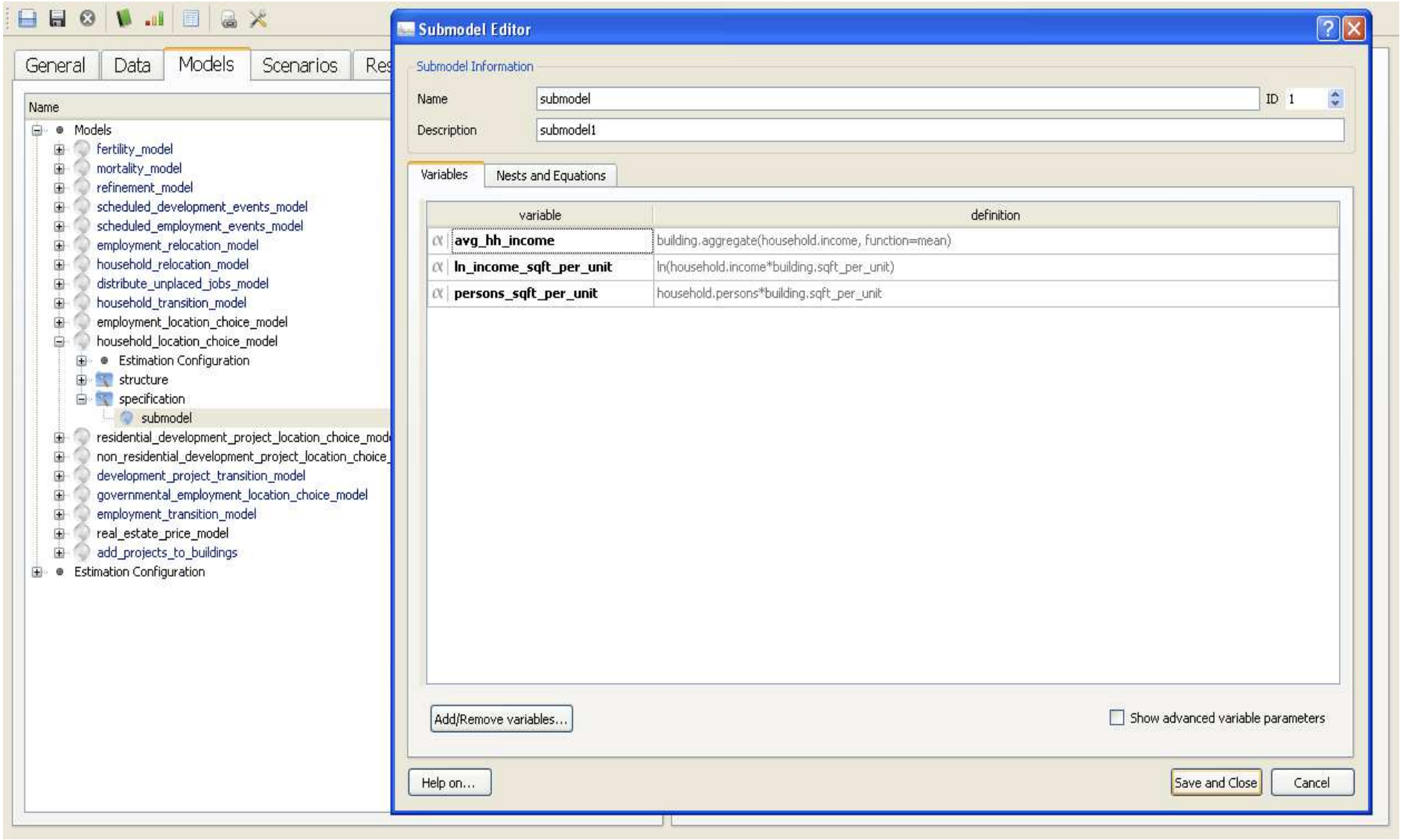
Tools to Manage Data



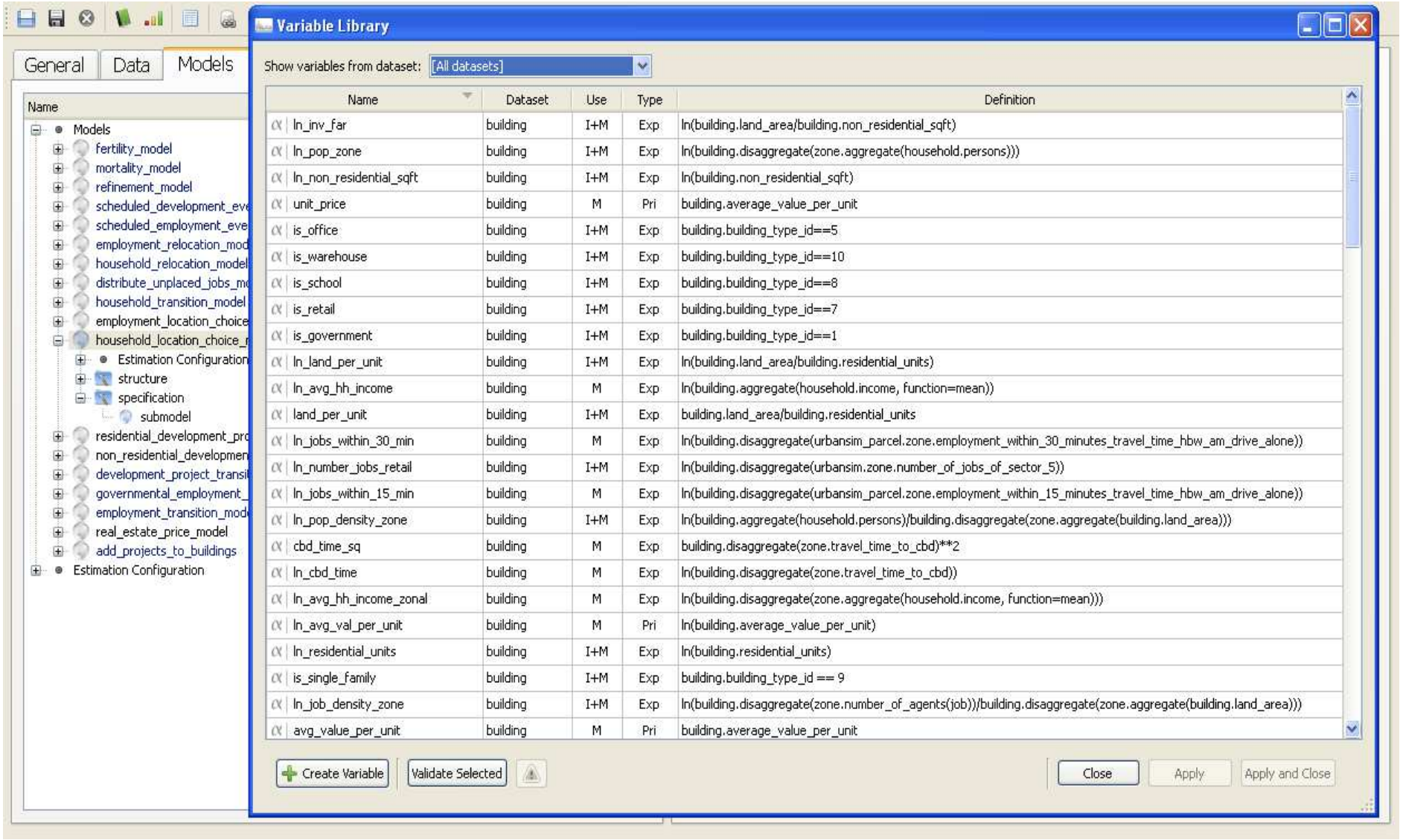
Database Connection (Importation, Exportation)



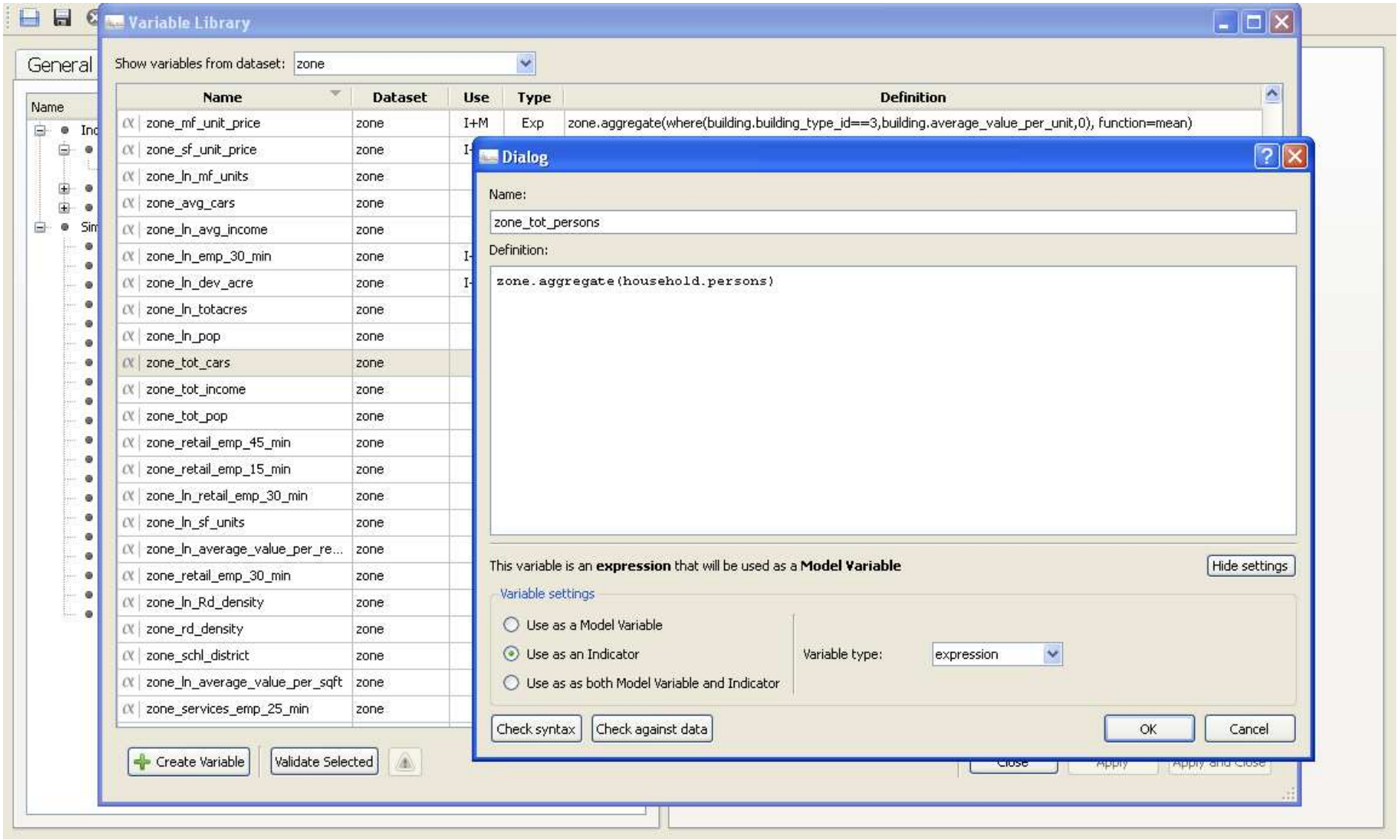
Models Definition



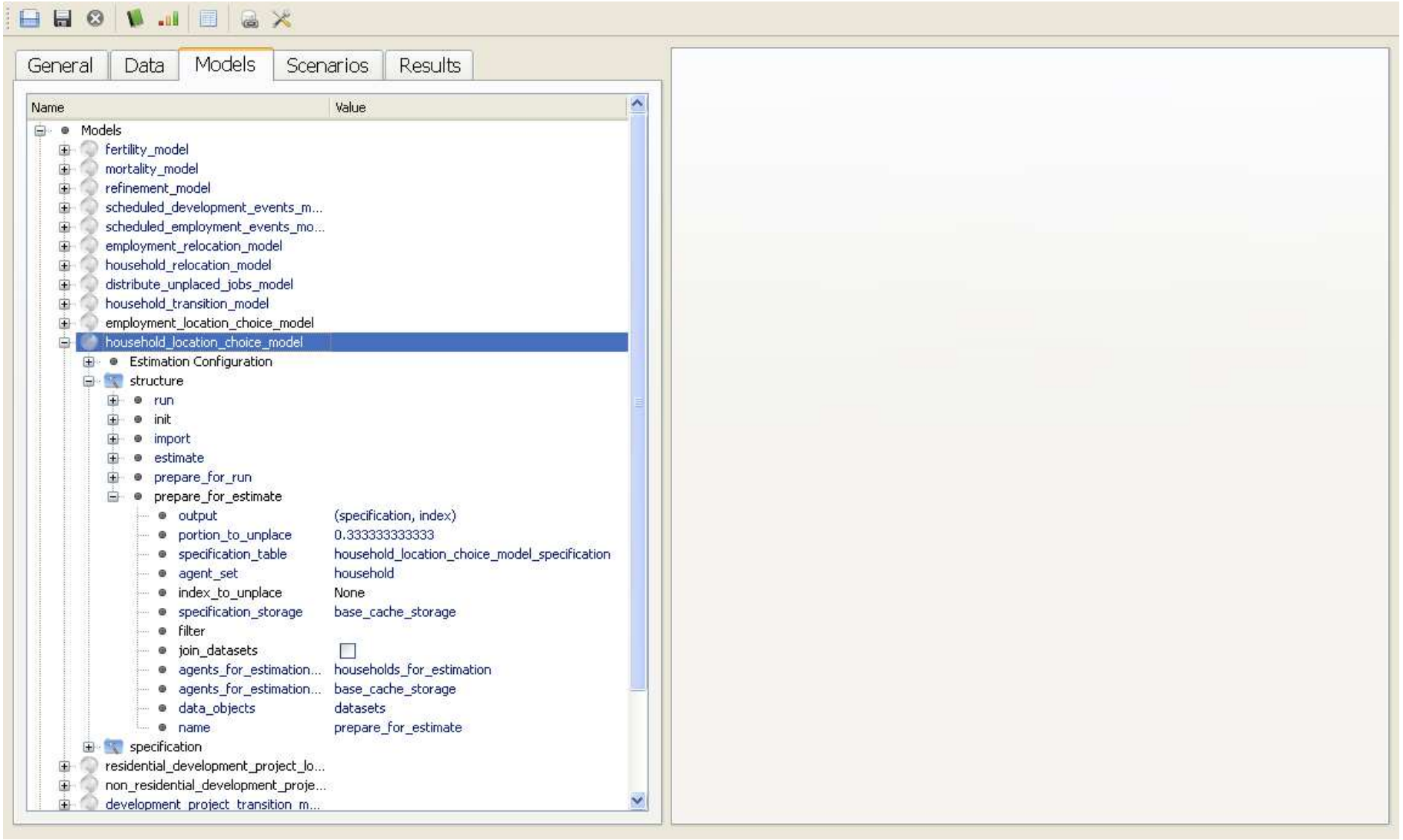
Variable Library



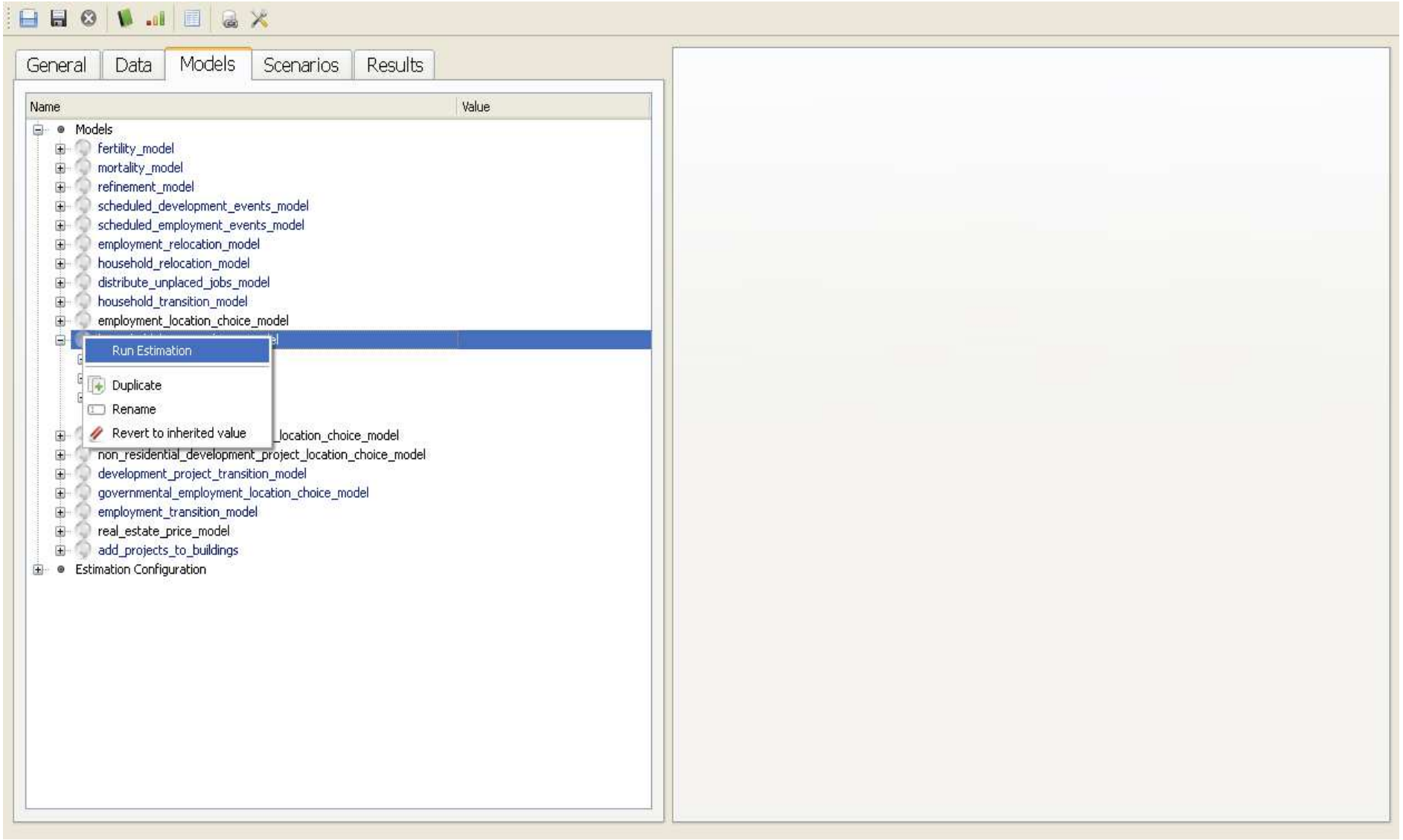
Variable (and/or Indicator) Creation



Model Parameters



Model Estimation

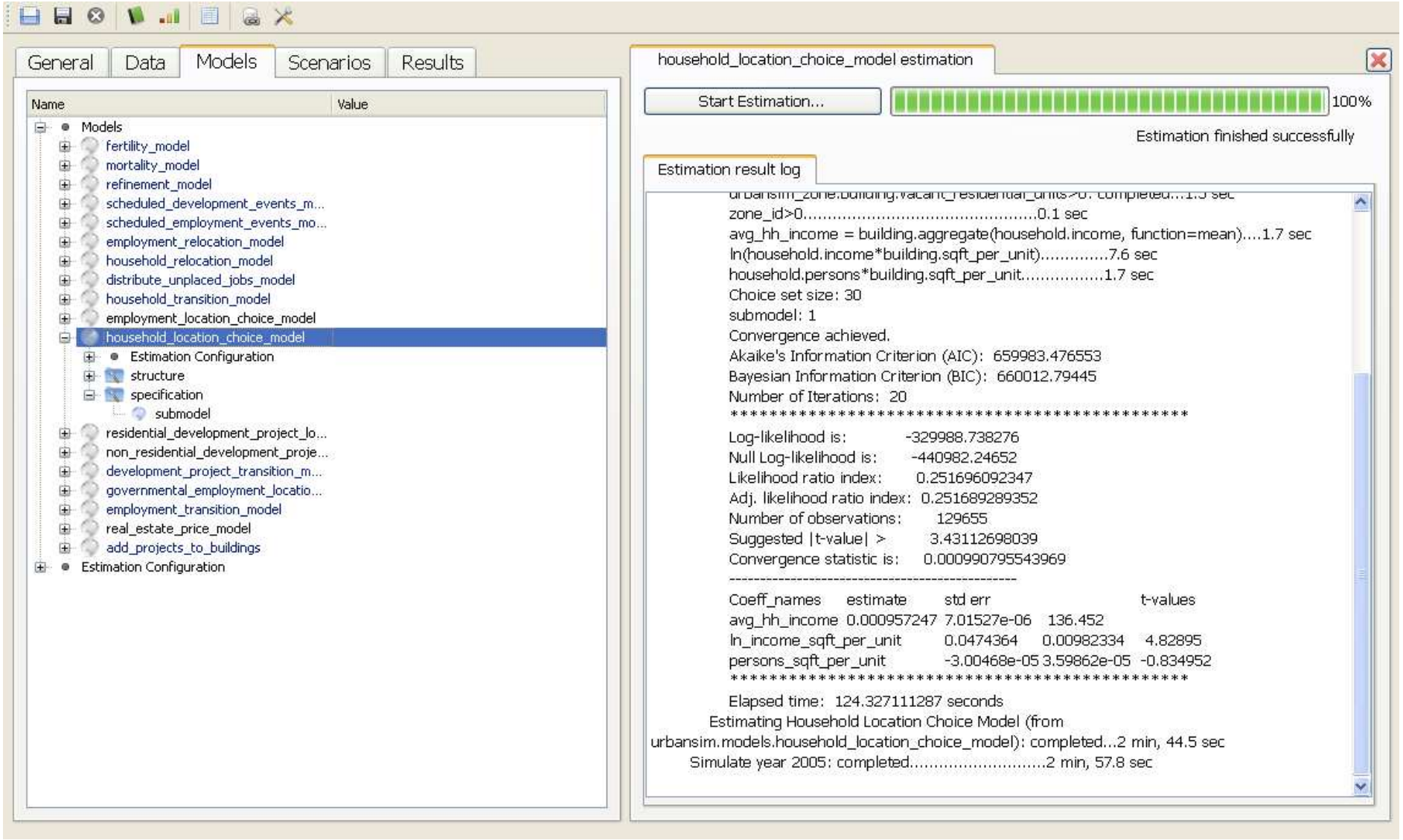


Model Estimation: Errors...

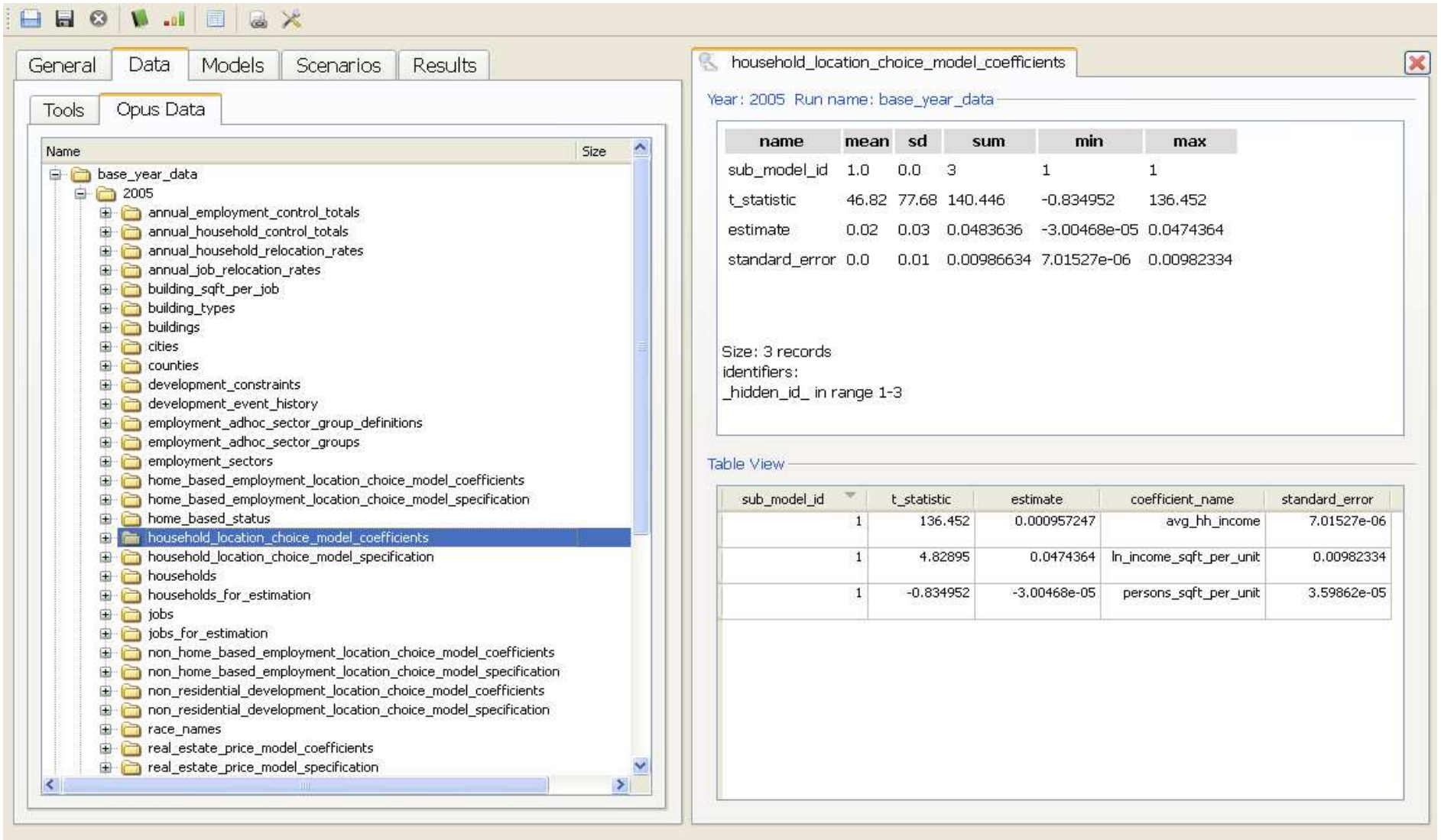
```
an)....2.7 sec
                cbd_time = building.disaggregate(zone.travel_time_to_cbd)....0.3
 sec
            Estimating Real Estate Price Model (from urbansim.models.real_estate
_price_model): completed...3.4 sec
        Simulate year 2005: completed.....................................4.0 sec

        Closing log file: C:\opus\data\san_antonio_zone/base_year_data\year_2005
_log.txt
    Starting simulation for year 2005: completed.........................4.0 sec

Start simulation run: completed.........................................4.0 sec

Closing log file: C:\opus\data\san_antonio_zone/base_year_data\run_model_system.
log
ERROR: Traceback (most recent call last):
  File "C:\opus\src\opus_gui\models_manager\run\run_estimation.py", line 124, in
 run
    self.er.estimate()
  File "C:\opus\src\urbansim\estimation\estimator.py", line 72, in estimate
    self.model_system.run(self.config, write_datasets_to_cache_at_end_of_year=Fa
lse)
  File "C:\opus\src\opus_core\model_coordinators\model_system.py", line 128, in
run
    write_datasets_to_cache_at_end_of_year=write_datasets_to_cache_at_end_of_yea
r)
  File "C:\opus\src\opus_core\model_coordinators\model_system.py", line 289, in
_run_year
    self.vardict[outputvar] = self.do_process(locals())
  File "C:\opus\src\opus_core\model_coordinators\model_system.py", line 370, in
do_process
    return eval(ev)
  File "<string>", line 1, in <module>
  File "C:\opus\src\opus_core\model.py", line 51, in logged_estimate_method
    results = estimate_method(*req_args, **opt_args)
  File "C:\opus\src\urbansim\models\real_estate_price_model.py", line 78, in est
imate
    estimate_config=estimate_config, debuglevel=debuglevel)
  File "C:\opus\src\opus_core\regression_model.py", line 226, in estimate
    dataset.compute_variables([outcome_attribute], dataset_pool=self.dataset_poo
l, resources=compute_resources)
  File "C:\opus\src\opus_core\datasets\abstract_dataset.py", line 625, in comput
e_variables
    (versions, value) = self.compute_variables_return_versions_and_final_value(n
ames, dataset_pool, resources, quiet)
  File "C:\opus\src\opus_core\datasets\abstract_dataset.py", line 644, in comput
e_variables_return_versions_and_final_value
    qualified_name = self.create_and_check_qualified_variable_name(name)
  File "C:\opus\src\opus_core\datasets\abstract_dataset.py", line 1804, in creat
e_and_check_qualified_variable_name
    self._check_dataset_name(vname.get_dataset_name())
  File "C:\opus\src\opus_core\datasets\abstract_dataset.py", line 1823, in _chec
k_dataset_name
    raise ValueError, 'different dataset names for variable and dataset'
ValueError: different dataset names for variable and dataset

Error returned from Estimation
Estimation Finished with sucess =  False
```

```
  File "C:\opus\src\opus_core\variables\variable.py", line 69, in logged_method
    results = compute_method(*req_args, **opt_args)
  File "C:\opus\src\opus_core\variables\variable.py", line 142, in compute_with_
dependencies
    self._solve_dependencies(dataset_pool)
  File "C:\opus\src\opus_core\variables\variable.py", line 206, in _solve_depend
encies
    (new_versions, value) = ds.compute_variables_return_versions_and_final_value
([(depvar_name, version)], dataset_pool)
  File "C:\opus\src\opus_core\datasets\abstract_dataset.py", line 652, in comput
e_variables_return_versions_and_final_value
    resources=resources, quiet=quiet, version=version))
  File "C:\opus\src\opus_core\datasets\abstract_dataset.py", line 1971, in _comp
ute_if_needed
    return self._compute_one_variable(variable_name, dataset_pool, resources=res
ources, quiet=quiet)
  File "C:\opus\src\opus_core\datasets\abstract_dataset.py", line 1915, in _comp
ute_one_variable
    data=variable.compute_with_dependencies(dataset_pool, compute_resources),
  File "C:\opus\src\opus_core\variables\variable.py", line 69, in logged_method
    results = compute_method(*req_args, **opt_args)
  File "C:\opus\src\opus_core\variables\variable.py", line 142, in compute_with_
dependencies
    self._solve_dependencies(dataset_pool)
  File "C:\opus\src\opus_core\variables\variable.py", line 206, in _solve_depend
encies
    (new_versions, value) = ds.compute_variables_return_versions_and_final_value
([(depvar_name, version)], dataset_pool)
  File "C:\opus\src\opus_core\datasets\abstract_dataset.py", line 652, in comput
e_variables_return_versions_and_final_value
    resources=resources, quiet=quiet, version=version))
  File "C:\opus\src\opus_core\datasets\abstract_dataset.py", line 1971, in _comp
ute_if_needed
    return self._compute_one_variable(variable_name, dataset_pool, resources=res
ources, quiet=quiet)
  File "C:\opus\src\opus_core\datasets\abstract_dataset.py", line 1903, in _comp
ute_one_variable
    index_name=id_name)
  File "C:\opus\src\opus_core\variables\variable_factory.py", line 81, in get_va
riable
    % (dataset_name, short_name))
LookupError: Incomplete variable specification for 'building.residential_units'
(missing package name).

Error returned from Estimation
Estimation Finished with sucess =  False
```

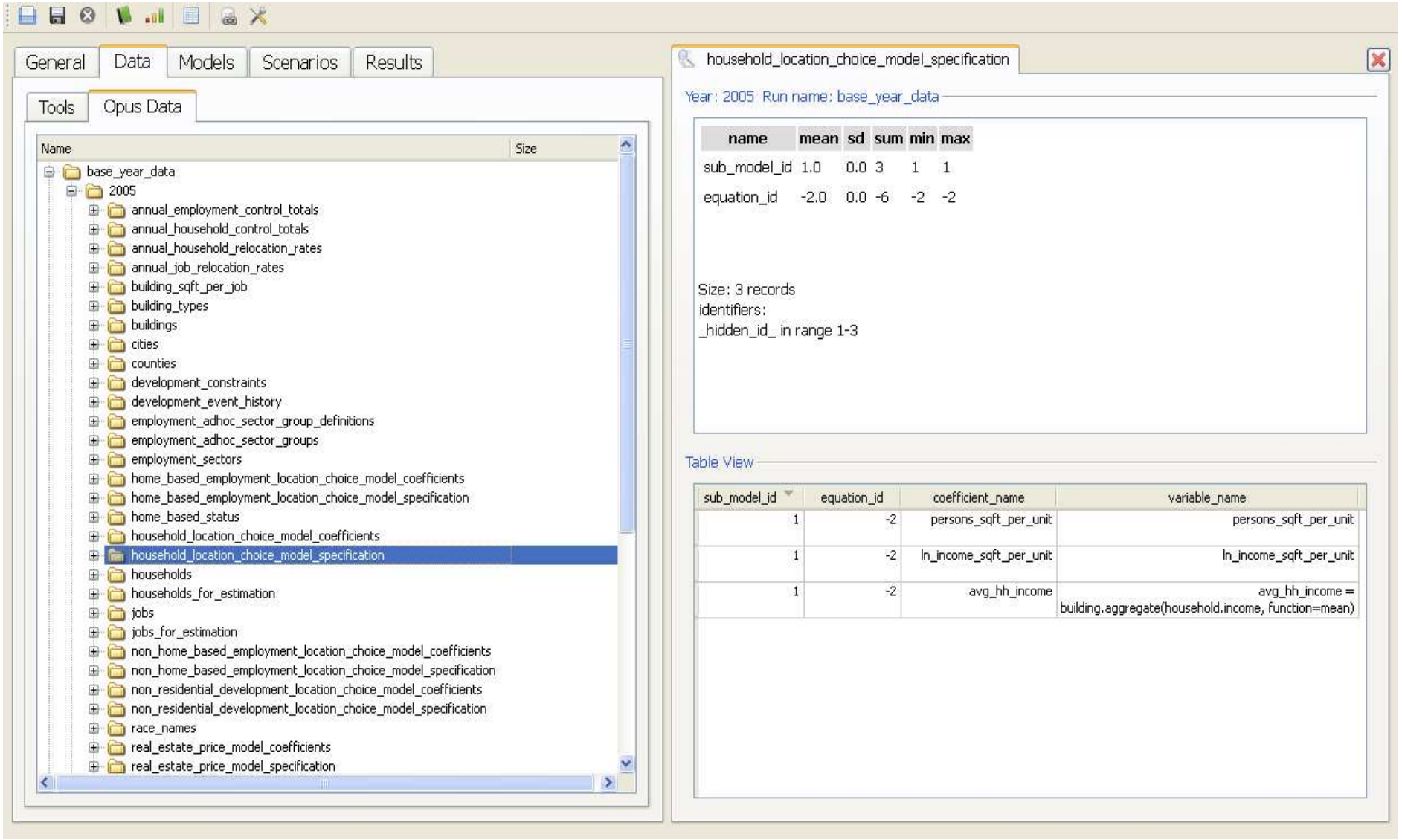
Model Estimation (Continued)



Model Coefficients



Model Specification



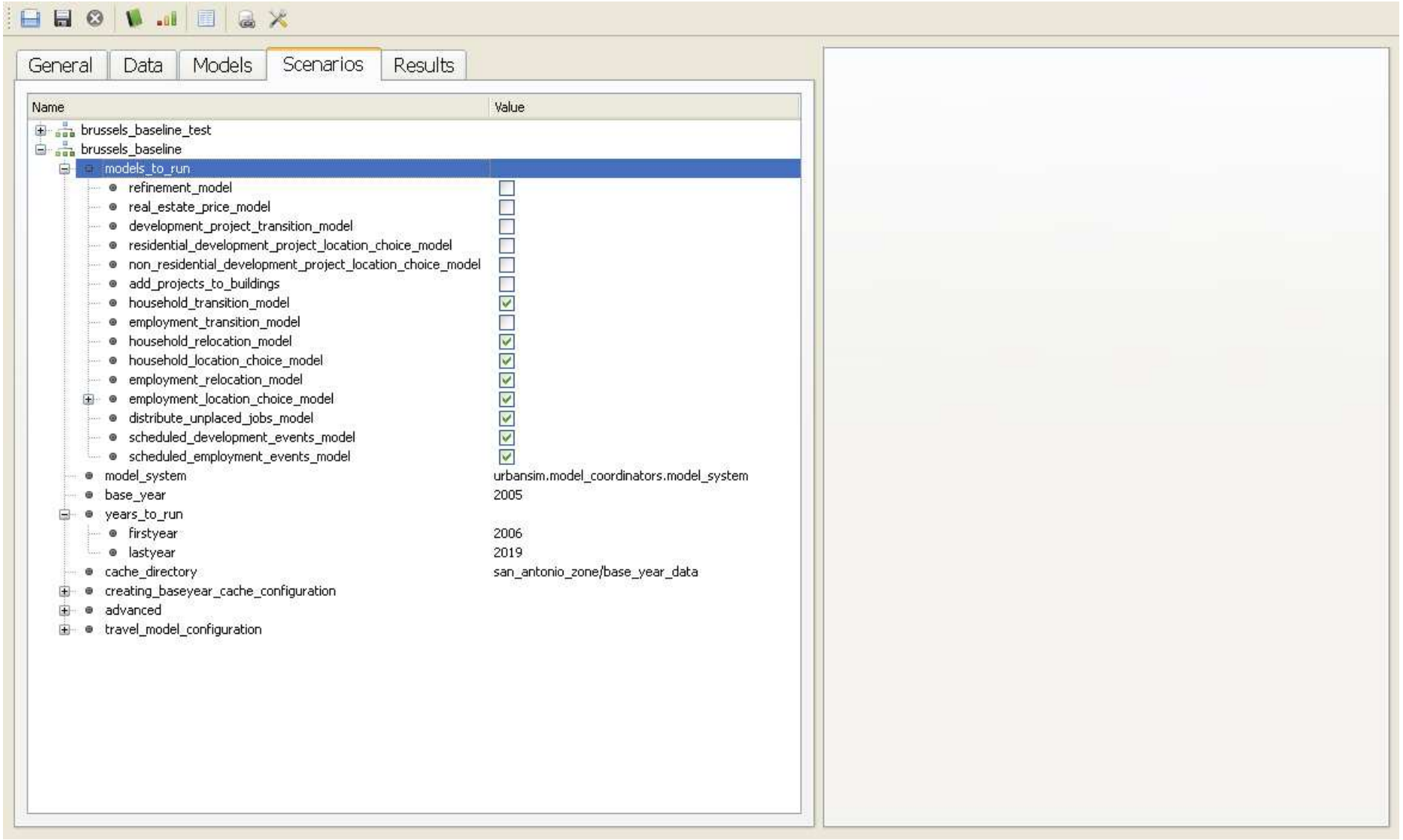
Model Estimation: Facts

- Special attention has to be paid on `xxx_id` (`zone_id`, `building_id`, `household_id`, etc.) variables: links between different tables, aggregations
- Error messages when missing data:
  - Some tables are mandatory for the zone version of UrbanSim: `www.urbansim.org`:
  - Some mandatory entries
- Specific `employment_sector_groups`: the submodels of the `employment_location_choice` model have to be adapted

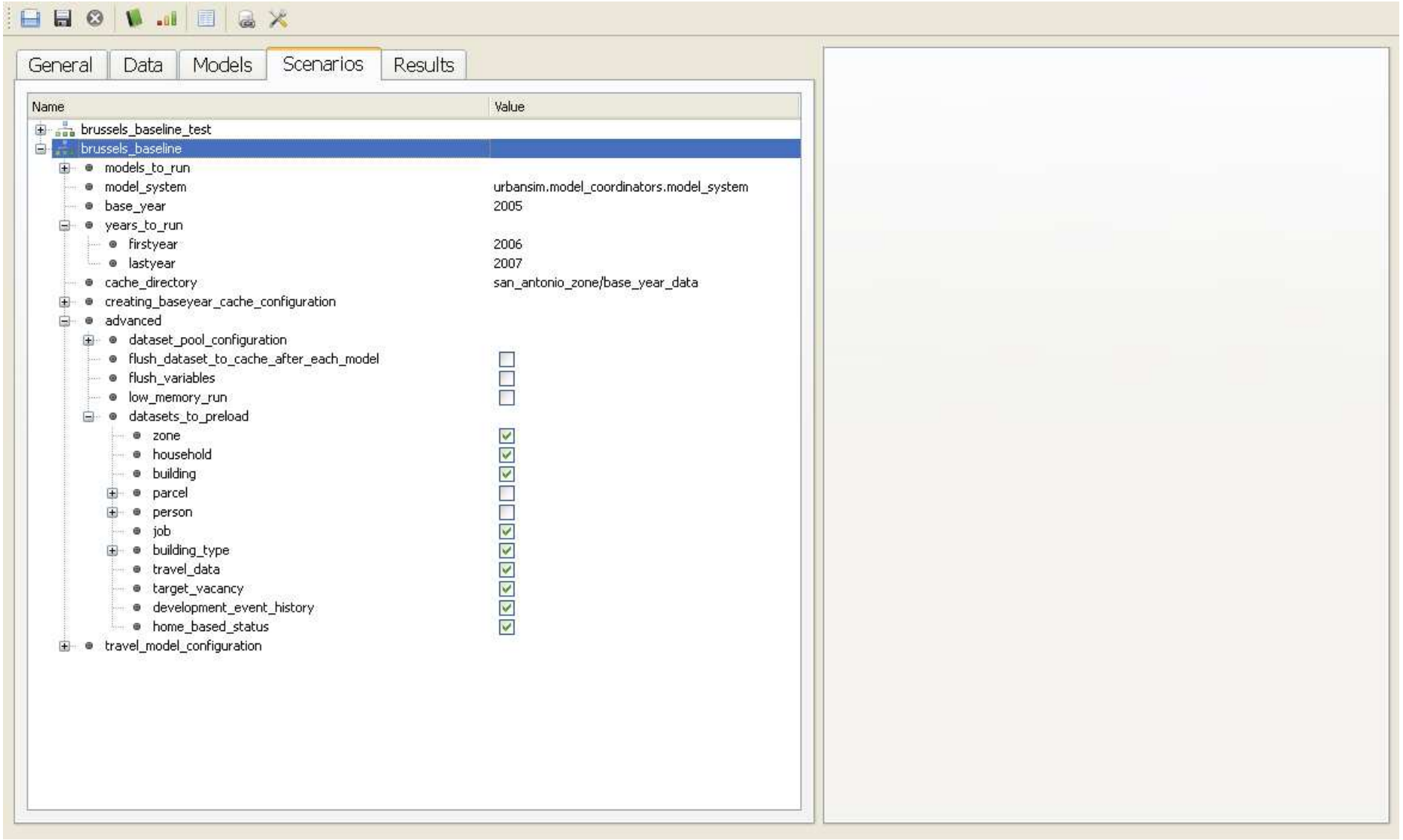
Model Estimation: Facts (Continued)

- To be able to estimate a submodel, your dataset must contain sufficient data from the considered type
- No capital letters in your table names and entries
- The formatting of your data is important, no entries with empty values
- Your dataset must contain the coefficients and specifications of the used models and some UrbanSim constants (here 14 additional tables):
  - `annual_household_relocation_rates` table for the `household_relocation` model
  - `annual_employment_relocation_rates` table for the `employment_relocation` model
  - ...

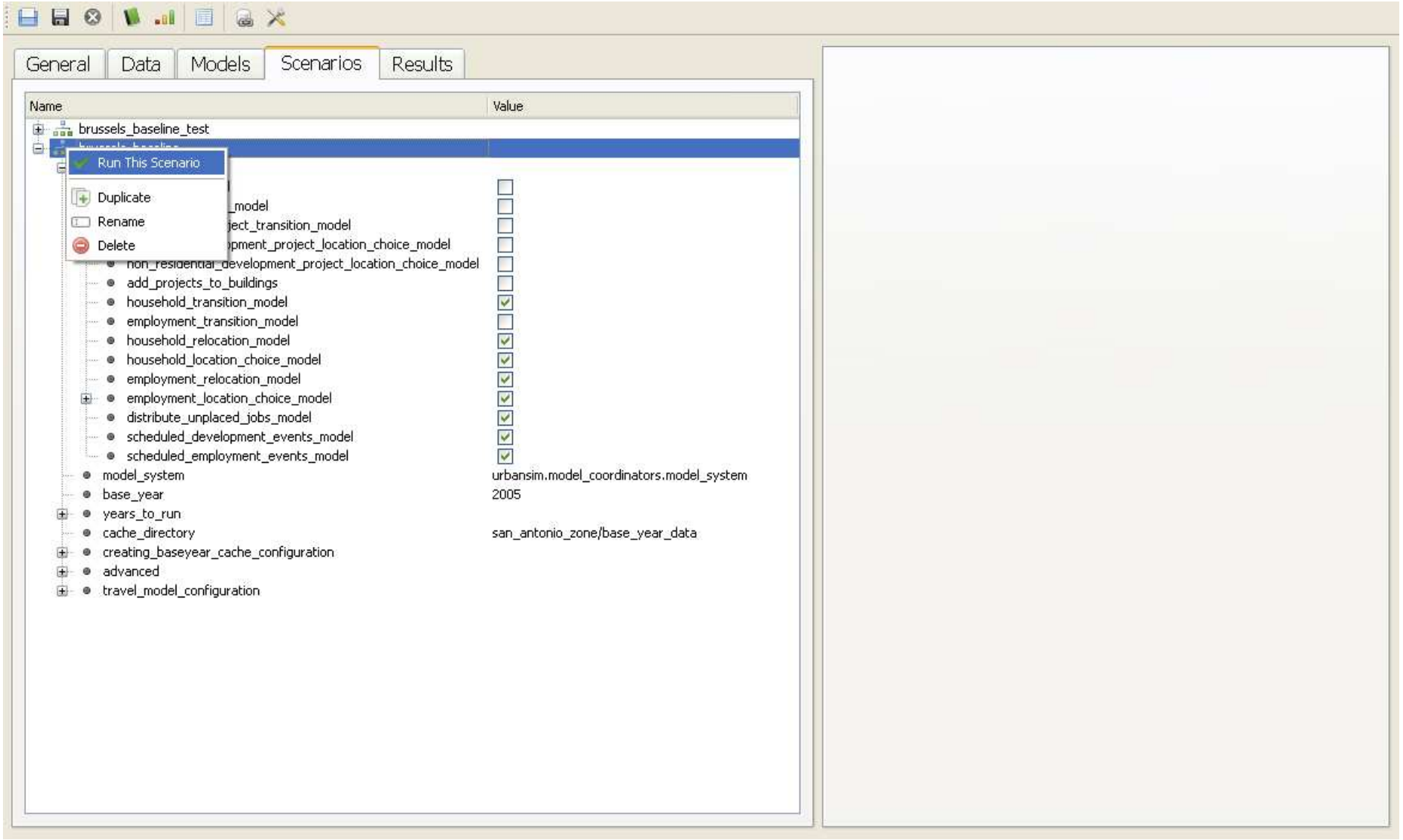
Model Simulation



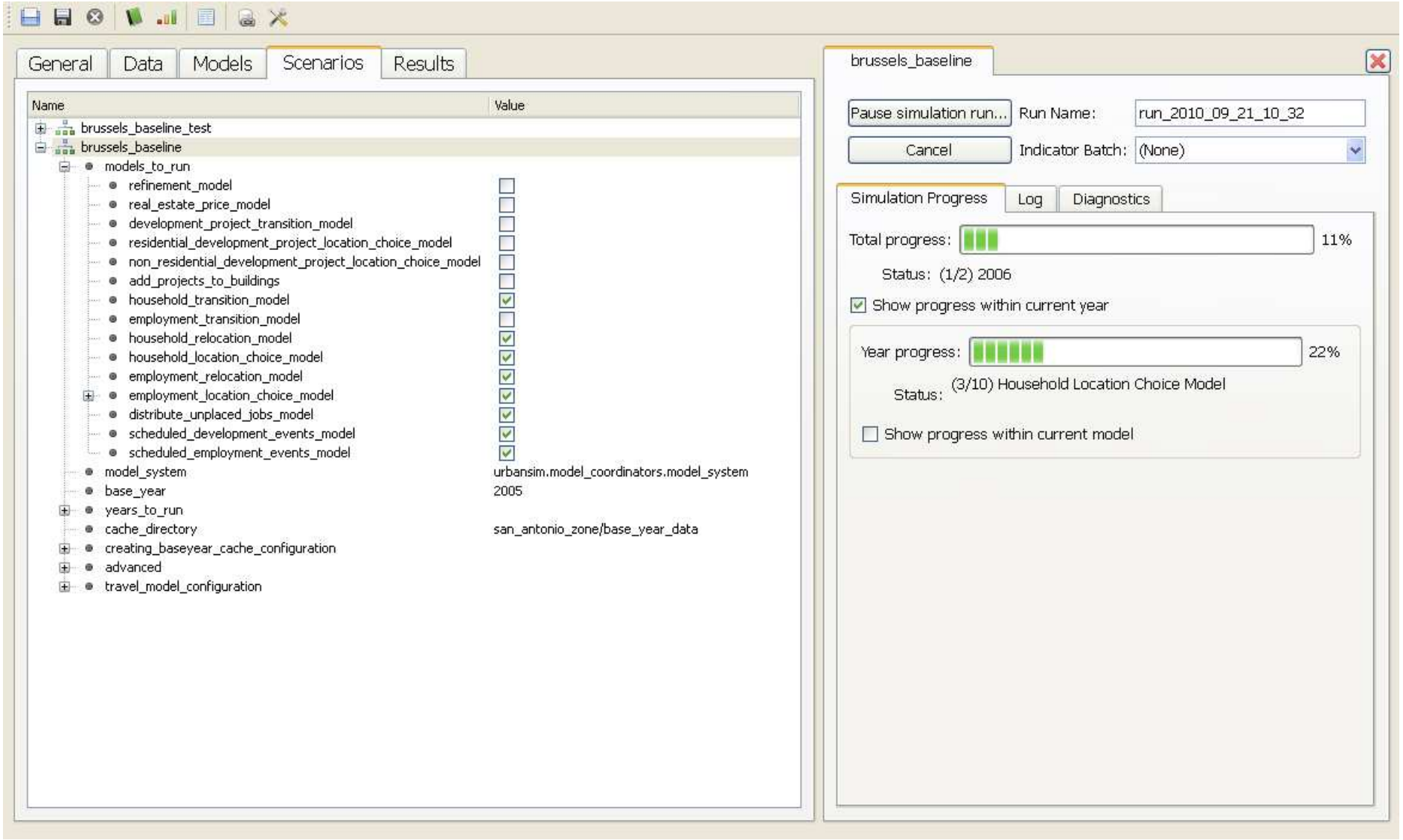
Model Simulation (Continued)



Model Simulation (Continued)



Model Simulation (Continued)



Model Simulation: Errors...

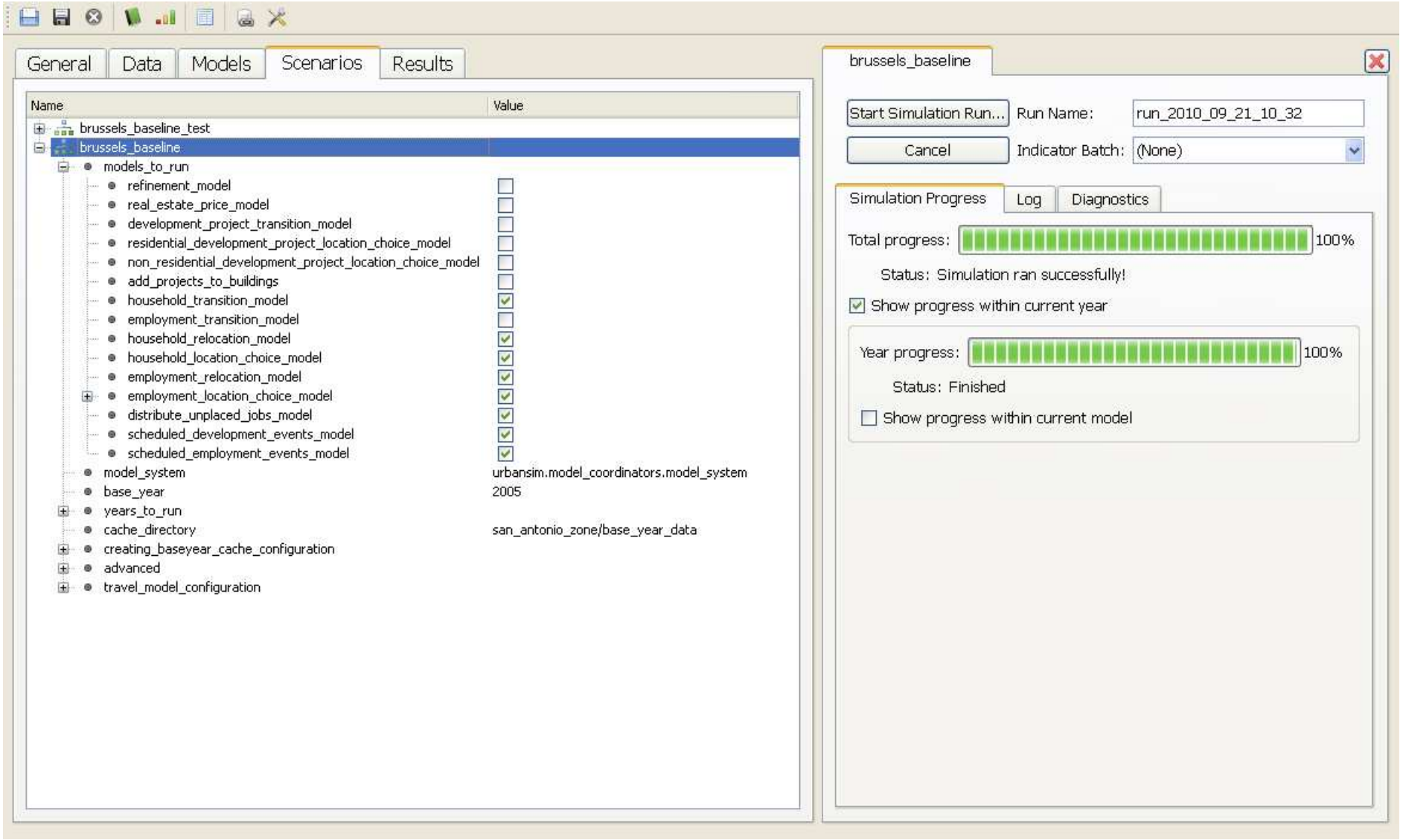
OPUS

```
    self.model_system.run(self.config, write_datasets_to_cache_at_end_of_year=Fa
lse)
  File "C:\opus\src\opus_core\model_coordinators\model_system.py", line 128, in
run
    write_datasets_to_cache_at_end_of_year=write_datasets_to_cache_at_end_of_yea
r)
  File "C:\opus\src\opus_core\model_coordinators\model_system.py", line 289, in
_run_year
    self.vardict[outputvar] = self.do_process(locals())
  File "C:\opus\src\opus_core\model_coordinators\model_system.py", line 370, in
do_process
    return eval(ev)
  File "<string>", line 1, in <module>
  File "C:\opus\src\opus_core\model.py", line 51, in logged_estimate_method
    results = estimate_method(*req_args, **opt_args)
  File "C:\opus\src\urbansim\models\agent_location_choice_model_member.py", line
 47, in estimate
    agents_index=agents_index[new_agents_index], **kwargs)
  File "C:\opus\src\urbansim\models\location_choice_model.py", line 198, in esti
mate
    debuglevel=debuglevel)
  File "C:\opus\src\opus_core\choice_model.py", line 364, in estimate
    self.create_interaction_datasets(agent_set, agents_index_for_estimation, est
imate_config, submodels=submodels)
  File "C:\opus\src\urbansim\models\location_choice_model.py", line 291, in crea
te_interaction_datasets
    nchunks=nchunks, chunksize=chunksize)
  File "C:\opus\src\opus_core\choice_model.py", line 511, in sample_alternatives
_by_chunk
    dataset_pool=self.dataset_pool
  File "C:\opus\src\opus_core\configurable.py", line 28, in config_run_method
    results = run_method(*req_args, **opt_args)
  File "C:\opus\src\opus_core\samplers\weighted_sampler.py", line 151, in run
    sampling_prob = column_stack([sampling_prob_for_chosen_choices, sampling_pro
b])
  File "C:\Python26\lib\site-packages\numpy\lib\shape_base.py", line 297, in col
umn_stack
    return _nx.concatenate(arrays,1)
MemoryError
```

Select OPUS

```
  File "C:\opus\src\opus_core\model_coordinators\model_system.py", line 289, in
_run_year
    self.vardict[outputvar] = self.do_process(locals())
  File "C:\opus\src\opus_core\model_coordinators\model_system.py", line 370, in
do_process
    return eval(ev)
  File "<string>", line 1, in <module>
NameError: name 'hrm_index' is not defined
Writing specification and coefficients into C:\opus\data\san_antonio_zone/base_y
ear_data: completed...2 min, 34.4 sec
ERROR: Traceback (most recent call last):
  File "C:\opus\src\opus_gui\scenarios_manager\run\run_simulation.py", line 206,
 in run
    run_id = run_manager.run_run(config, run_name = run_name)
  File "C:\opus\src\opus_core\services\run_server\run_manager.py", line 107, in
run_run
    model_system.run_multiprocess(run_resources)
  File "C:\opus\src\opus_core\model_coordinators\model_system.py", line 466, in
run_multiprocess
    self._run_each_year_as_separate_process(start_year, end_year, seed_array, re
sources)
  File "C:\opus\src\urbansim\model_coordinators\model_system.py", line 35, in _r
un_each_year_as_separate_process
    'urbansim.model_coordinators.model_system', resources, optional_args=['--log
-file-name', log_file_name])
  File "C:\opus\src\opus_core\model_coordinators\model_system.py", line 563, in
_fork_new_process
    self.forked_processes[-1].wait()
  File "C:\opus\src\opus_core\fork_process.py", line 79, in wait
    self.check_status()
  File "C:\opus\src\opus_core\fork_process.py", line 85, in check_status
    raise StandardError("Child python process exited with failure.\nCalling modu
le: %s\nSystem command: %s" % (self.module_name, self.python_cmd))
StandardError: Child python process exited with failure.
Calling module: urbansim.model_coordinators.model_system
System command: ['C:\\Python26\\python.exe', 'C:\\opus\\src\\urbansim\\model_coo
rdinators\\model_system.py', '-r', 'c:\\docume~1\\admini~1\\locals~1\\temp\\tmpa
mktgl\\resources.pickle', '--log-file-name', 'run_multiprocess.log']

Error returned from Model
Traceback (most recent call last):
  File "C:\opus\src\opus_gui\scenarios_manager\run\run_simulation.py", line 34,
in run
    self.modelguielement.model.run()
  File "C:\opus\src\opus_gui\scenarios_manager\run\run_simulation.py", line 227,
 in run
    self.finishedCallback(succeeded, run_name = run_name)
  File "C:\opus\src\opus_gui\scenarios_manager\run\run_simulation.py", line 77,
in finishedCallback
    run_name = run_name)
  File "C:\opus\src\opus_gui\results_manager\results_manager_functions.py", line
 62, in delete_simulation_run
    get_manager_instance('results_manager').delete_run(run_node)
  File "C:\opus\src\opus_gui\results_manager\results_manager.py", line 75, in de
lete_run
    self.xml_controller.delete_run(run_node, force=force)
  File "C:\opus\src\opus_gui\results_manager\controllers\xml_configuration\xml_c
ontroller_results.py", line 67, in delete_run
    cache_directory = run_node.find('cache_directory').text
AttributeError: 'NoneType' object has no attribute 'find'
```

Model Simulation (Continued)

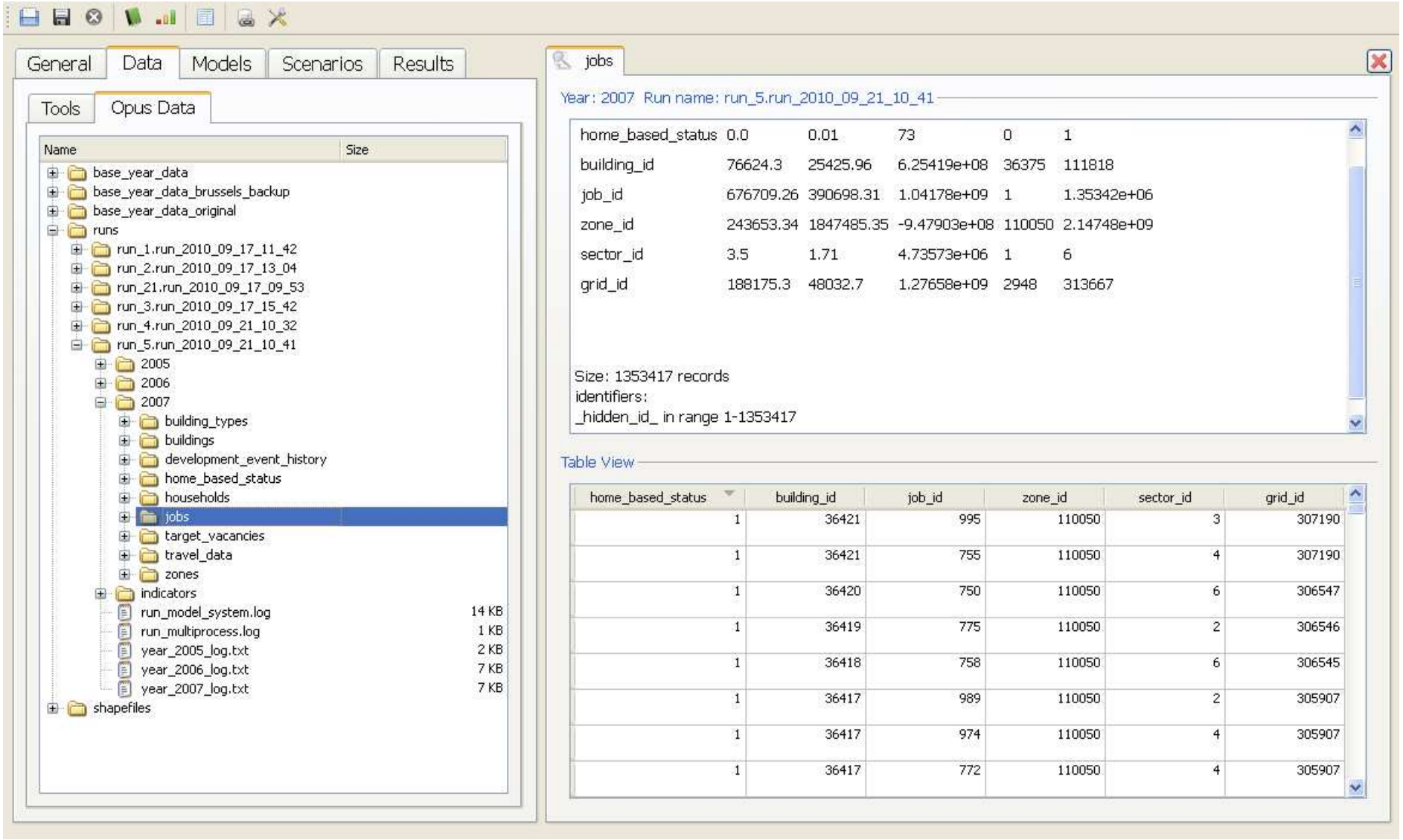


Model Simulation: Facts

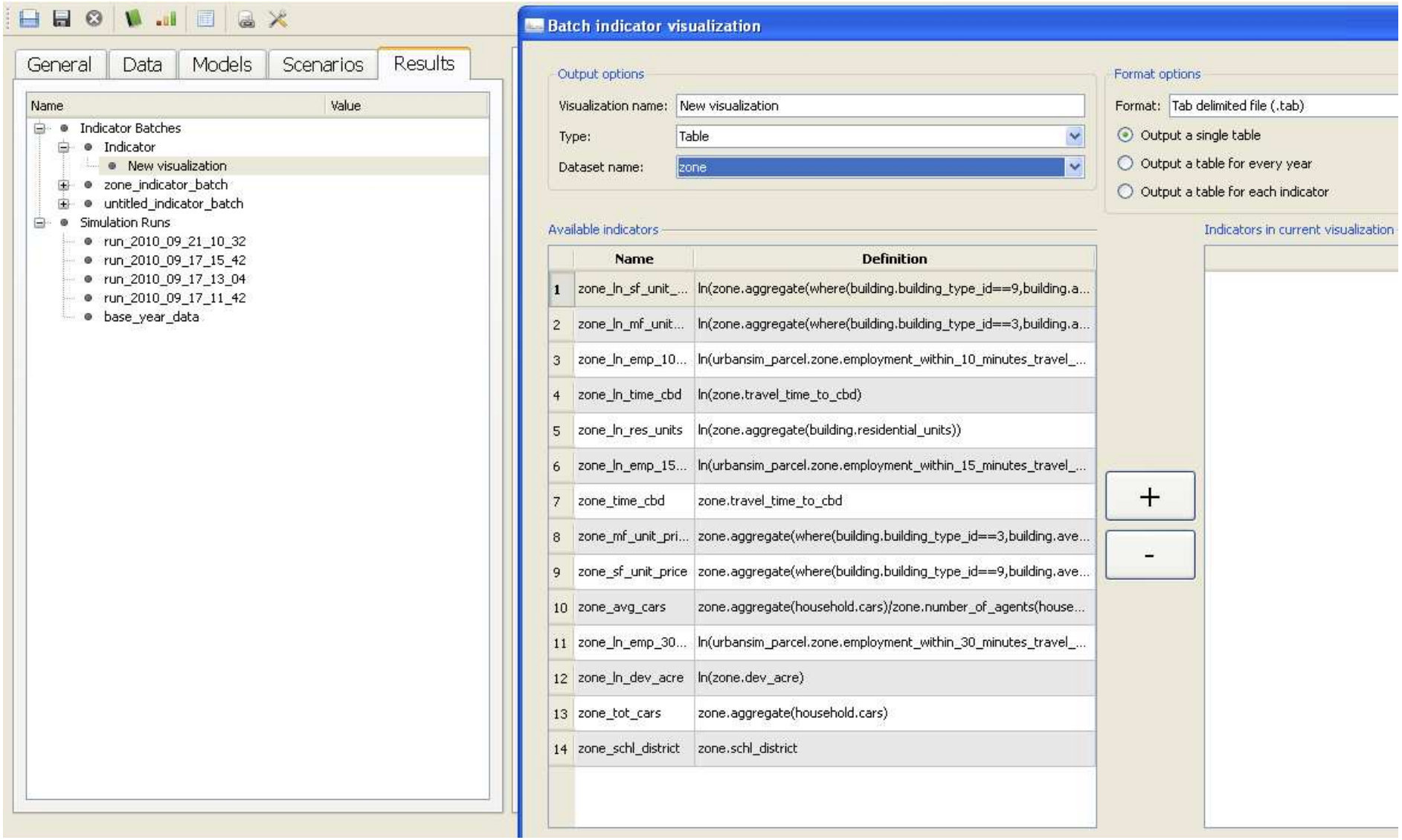
- Some models have to be run together (*i.e.* the `household_relocation` model has to be run before the `household_location_choice` model)
- Some models cannot be deleted because mandatory
- In the current version, control totals are used:
  - `annual_household_control_totals`
  - `annual_employment_control_totals`

  but fertility and mortality models can be used in addition to that.

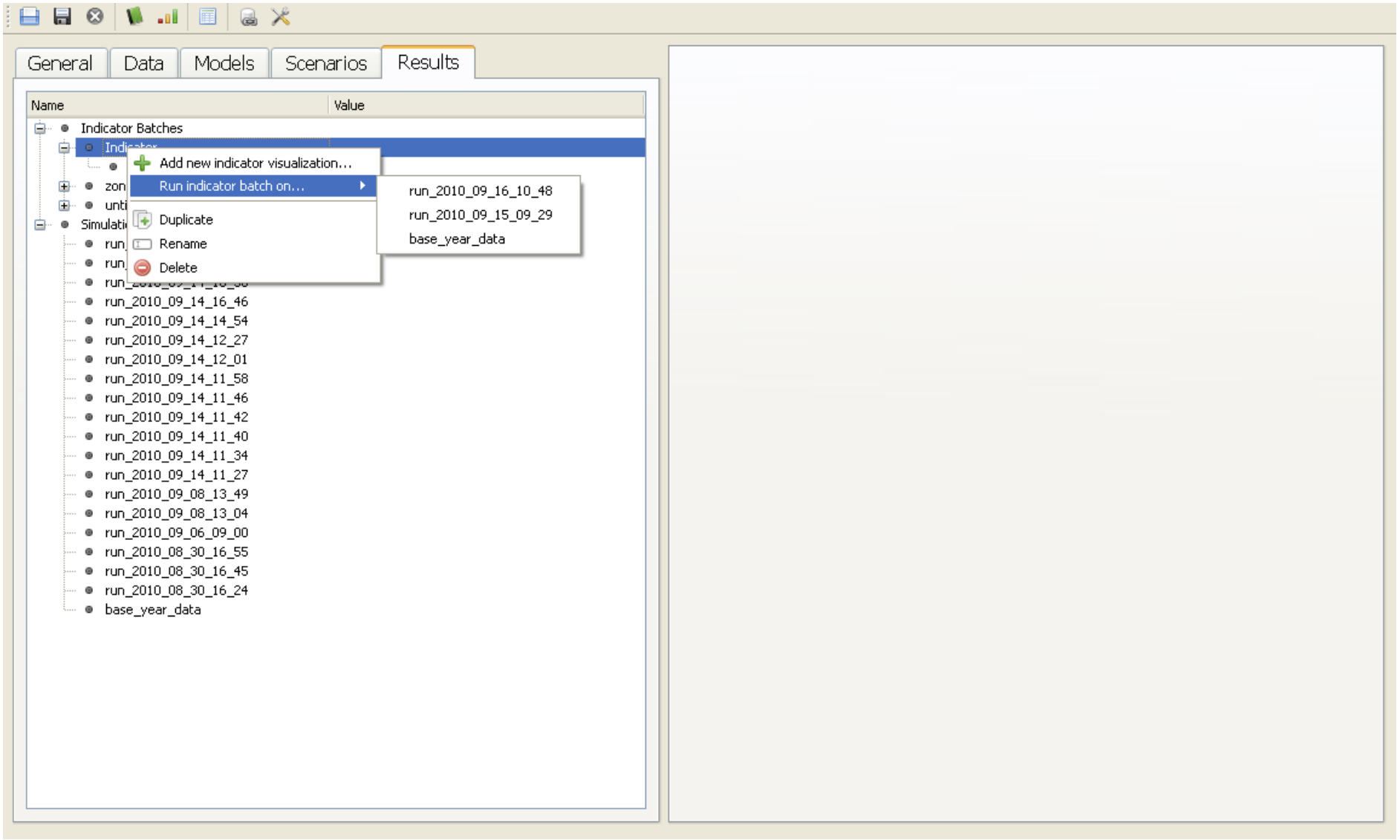
Simulation Results: Data



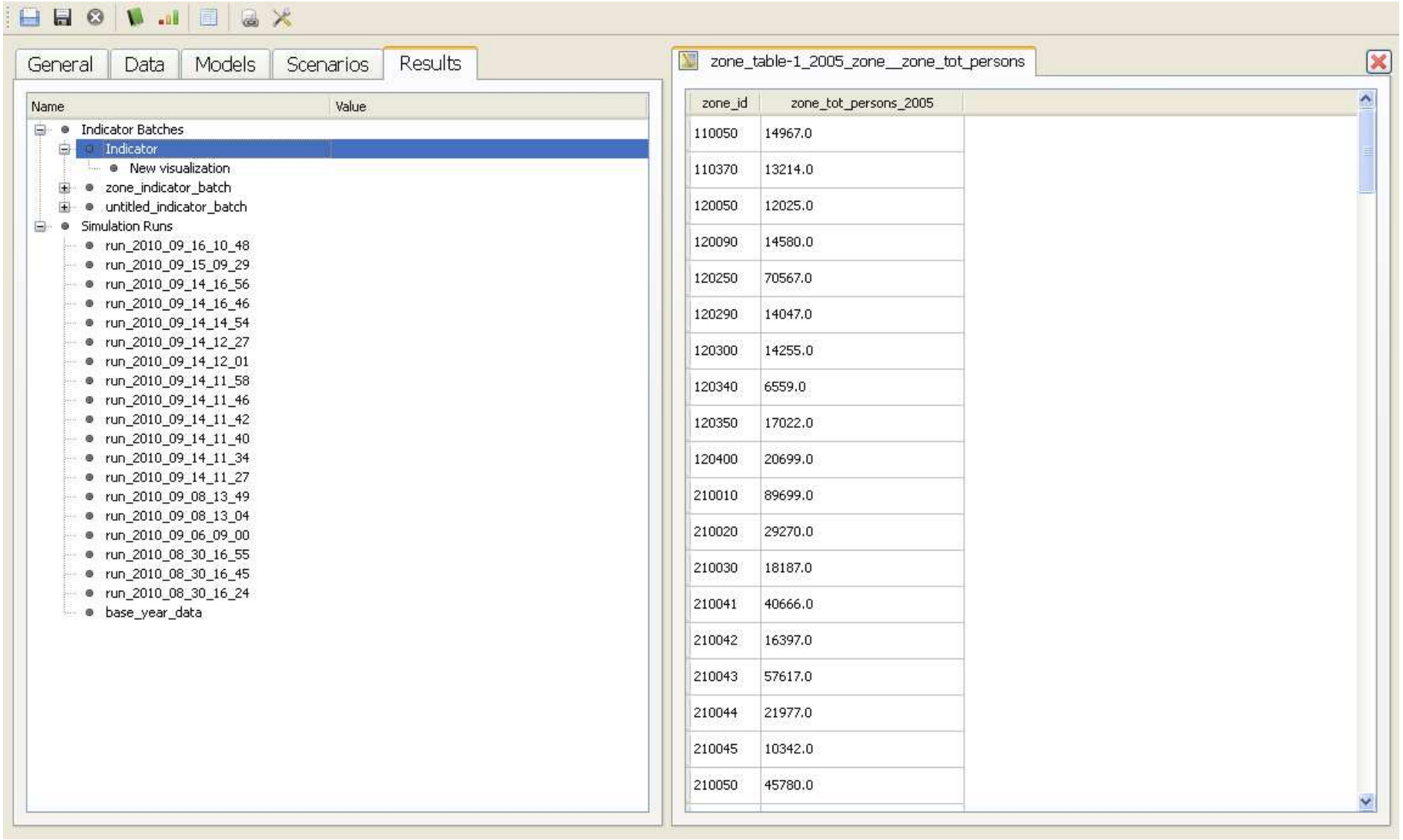
Simulation Results: Indicators



Simulation Results: Indicators (Continued)



Simulation Results: Indicators (Continued)



zone_id	zone_tot_persons_2005
110050	14967.0
110370	13214.0
120050	12025.0
120090	14580.0
120250	70567.0
120290	14047.0
120300	14255.0
120340	6559.0
120350	17022.0
120400	20699.0
210010	89699.0
210020	29270.0
210030	18187.0
210041	40666.0
210042	16397.0
210043	57617.0
210044	21977.0
210045	10342.0
210050	45780.0

Simulation Results

zone_table-1_2005_zone__zone_tot_persons

zone_id	zone_tot_persons_2005
110050	14967.0
110370	13214.0
120050	12025.0
120090	14580.0
120250	70567.0
120290	14047.0
120300	14255.0
120340	6559.0
120350	17022.0
120400	20699.0
210010	89699.0
210020	29270.0
210030	18187.0
210041	40666.0
210042	16397.0
210043	57617.0
210044	21977.0
210045	10342.0
210050	45780.0

zone_table-1_2006_zone__zone_tot_persons

zone_id	zone_tot_persons_2006
110050	14256.0
110370	13208.0
120050	13256.0
120090	14788.0
120250	70211.0
120290	14910.0
120300	15034.0
120340	6807.0
120350	18071.0
120400	20356.0
210010	85418.0
210020	28566.0
210030	17332.0
210041	37758.0
210042	15329.0
210043	54006.0
210044	21328.0
210045	10045.0
210050	42830.0

zone_table-1_2007_zone__zone_tot_persons

zone_id	zone_tot_persons_2007
110050	14316.0
110370	13570.0
120050	13504.0
120090	14947.0
120250	70846.0
120290	15150.0
120300	15337.0
120340	6944.0
120350	18215.0
120400	20536.0
210010	85613.0
210020	28693.0
210030	17325.0
210041	37772.0
210042	15344.0
210043	54217.0
210044	21524.0
210045	10101.0
210050	42825.0

Simulation Results (Continued)

zone_table-1_2005_zone__zone_tot_cars

zone_id	zone_tot_cars_2005
110050	4076.0
110370	3698.0
120050	3407.0
120090	3950.0
120250	20758.0
120290	3852.0
120300	3891.0
120340	1880.0
120350	4841.0
120400	5637.0
210010	25212.0
210020	8932.0
210030	5269.0
210041	12281.0
210042	5275.0
210043	16941.0
210044	6200.0
210045	3644.0
210050	14335.0

zone_table-1_2006_zone__zone_tot_cars

zone_id	zone_tot_cars_2006
110050	3960.0
110370	3737.0
120050	3695.0
120090	4038.0
120250	20809.0
120290	4021.0
120300	4054.0
120340	1938.0
120350	5103.0
120400	5584.0
210010	24590.0
210020	8877.0
210030	5132.0
210041	11792.0
210042	5076.0
210043	16331.0
210044	6115.0
210045	3607.0
210050	13846.0

zone_table-1_2007_zone__zone_tot_cars

zone_id	zone_tot_cars_2007
110050	3977.0
110370	3814.0
120050	3765.0
120090	4070.0
120250	20935.0
120290	4140.0
120300	4122.0
120340	1977.0
120350	5138.0
120400	5652.0
210010	24631.0
210020	8896.0
210030	5122.0
210041	11793.0
210042	5081.0
210043	16365.0
210044	6173.0
210045	3594.0
210050	13856.0

Models that Work

- `household_transition` model
- `employment_transition` model
- `household_relocation` model
- `household_location_choice` model
- `employment_relocation` model
- `employment_location_choice` model
- `distribute_unplaced_jobs` model
- `scheduled_development_events` model
- `scheduled_employment_events` model

What Still Does Not Work

Development Project Location Choice Model:

- Dependent on your specific `building_types`
- Lack of data...
- Dataset used for estimation: `development_event_history`
- Must have sufficient data for each submodel to be estimated
- `residential_units_capacity` $>$ `residential_units`

What Still Does Not Work (Continued)

Real Estate Price Model:

- Different `building_types` dataset
- Specific submodel for each `building_type`
- Tried to simplify or remove some submodels to cope to the Brussels case and to the specific `building_types`
- However, some specificities of the San Antonio project remain...
- Dataset concerned here for estimation: `buildings` and `building_types`
- Outcome of the model: `building.average_value_per_unit`
- The same kind of problem must have appeared in the other case studies (Zurich, Paris)

What Remains to Do with the New Data

What Remains to Do with the New Data:

- Results visualization
- Integration of Matsim add-on: more precise transportation simulation aspects
  - Python $\leftrightarrow$ Java
  - Communication
  - Data transfer