February 2022

Activity-Based Model: Frequently Asked Questions

Coordinated Travel – Regional Activity-Based Modeling Platform (CT-RAMP) for Atlanta Regional Commission

Atlanta Regional Commission
229 Peachtree Street NE, Suite 100
Atlanta, Georgia 30303

GENERAL

Is the ABM now considered the "Official Travel Demand Model of the ARC"?

Yes, with the ARC Board's Regional Transportation Plan adoption, the ABM is now the official travel demand model.

If so, I imagine the only ones using the older 4-step ARC model are on those studies that began prior to release of the ABM… correct?

Yes, correct.

Is a Cloud model available for the ARC ABM?

ARC and Citilabs are working together to develop a Cube Cloud version. Once it is setup and fully tested, details of the Cloud model will be available.

INSTALLATION & OPERATION

What are the minimum and preferable computer requirements for running the ARC ABM?

- Recommended number of cores is at least 12 physical cores.
- Recommended available hard drive space at the beginning of a run is at least 120 GB.
- Equipped with at least 128 GB of RAM and running 64-bit Windows OS.
- Installed with 64-bit Java Development Kit 1.7 or later.
- Installed with Cube Voyager 6.1 or later, with at least one installation with at least a seat Cube Cluster license.
- Installed with VoyagerFileAPIInstaller.msi, which places 64-bit Voyager matrix access API files under the *Citilabs\VoyagerFileAPI* directory.
- Include *C:\Program Files\Citilabs\VoyagerFileAPI* in the system %PATH%.
- Installed with Microsoft Visual C++ 2012 (x64) redistributable package (downloadable from *http://www.microsoft.com/enus/download/details.aspx?id=30679*)

What are the specs of ARC servers and how many cluster nodes are used?

The current version of the ARC ABM is run on one Microsoft Windows Server machine. The following table shows the specifications of the servers. The total number of extra cluster nodes ARC uses is 63 or 47 depending on the server used.

Alternatively, the ARC ABM can be run using multiple servers. In the multiple server setup, the total number of extra cluster nodes is set at the sum of the physical cores of the servers used less one.

Server	Physical Cores	Logical Cores	CPU	RAM	HD Capacity
Main/Node 1	64	128	Intel Xeon Gold 6130 @ 2.10 GHz	512 GB	7.9 TB
Node 2	48	96	Intel Xeon E5-4657L v2 @ 2.40GHz	448 GB	9.3 TB
Node 3	48	96	Intel Xeon E5-4657L v2 @ 2.40GHz	448 GB	1.6 TB

What version of Cube is compatible with the ARC ABM?

ARC ABM is currently compatible with Cube version 6.4.5.

How much does the hardware cost to run the ARC ABM?

The cost varies substantially based on the computer specifications. The minimum specification of 128 GB RAM and 24 processors probably retails off-the-shelf between $7,000 and $10,000 depending on other accessories.

What software and version is required to run the ARC PopSyn? How much does it cost? Where can I get the software?

In 2019, ARC revised the PopSyn so that it can be run in MySQL instead of MS SQL. This revision benefits the users who can't afford MS SQL programs. To run the revised PopSyn, the users need to install MySQL and a specific version of JAVA on their machine. Details about hardware and software requirements are provided in a ReadMe.txt file with the PopSyn software.

What are the minimum computer requirements for running the ARC PopSyn?

The ARC PopSyn is not run in a distributed mode, so a single core will work. The maximum RAM specified in the PopSyn call is 15GB. 24GB to 32GB of memory is recommended.

The file "accessibilities.properties" contains property settings for the accessibility calculator. Is this required for model run?

This file is used to post-process the completed model run to generate the logsumbased accessibilities. It is not an input file to the model.

How is the ARC ABM setup on multiple computers? Do all machines need to have cluster installed? Can the ARC ABM be setup on a single computer?

The model root folder on the main machine (C:\ARCABM) is mapped with the same network drive (e.g., drive "T") across all machines. In each machine, cluster nodes are launched by accessing the mapped drive. For details, refer to ARC's online Activity-Based Model User Guide. The ARC ABM is usually setup on a single machine. In this case, mapping the drive is unnecessary, and the model is setup and run under C:\ARCABM\{scenario folder}.

We have a different version of JAVA installed on our machines. Does other version work with the ARC ABM? What settings in the ABM need to be changed?

ARC ABM has been tested with Java Platform Development Kit (JDK) 7u45 and 8u91. The Java version is specified in the following parts:

- runMain.cmd, runNode1.cmd, runNode2.cmd, runNode3.cmd
- System PATH variable
- Model main script
  ;jdk64:{jdk64,editbox,"Directory to 64bit JDK",N,"C:/Progra~1/Java/jdk1.8.0_91"}

We currently have several workstations with cluster that we use for modeling. They don't have the RAM recommended in the online User's Guide, but we are looking into making that upgrade. Can the model be executed successfully on computers with less than 128 GB of RAM? Does the RAM primarily affect the runtime or is it required for processes to execute without crashing? If we run the model on computers without the recommended configuration, would it be better to run the model across multiple machines/nodes instead of on just one computer/node?

it is better to use the recommended specs, so 128 GB of RAM, from a model run times perspective is preferable on a single, high-performing machine, as opposed to running the ARC ABM across multiple machines.

The online documentation says to install 64-bit Java Development Kit (JDK) 1.8 or later. My understanding is that Oracle is the current distributor of Java, but their website doesn't seem to use this naming convention. The Oracle website has a download for Java SE Development Kit 8u301? Is this the same as JDK 1.8? If not, where does one get JDK 1.8?

Go to https://www.oracle.com/java/technologies/downloads/#java8-windows, then look for the download link associated with jdk-8u301-windows-x64.exe.

The online User's Guide mentions installing VoyagerFileAPInstaller.msi. Do you have any recommendations on how to get the VoyagerFileAPInstaller.msi? Is this something we have to request from Bentley?

No. Please contact the ARC modeling group for the VoyagerFileAPIlstaller.zip file.

If we run on multiple computers, the 2017 PDF User's Guide indicates that we need to install the ARCTourBasedModel_Node.zip file on each node. This file isn't available on the model download page. Is this file still necessary?

Yes. Please contact the ARC modeling group for the ARCTourBasedModel_Node.zip file. The difference with a multiple server setup with ARCTourBasedModel_Node.zip compared to a one server setup appears in the config folder. For a one server setup, these files are needed: jppf-node1.properties, log4j-node1.xml, and runNode1.cmd. For a two server setup, these files are also needed: jppf-node2.properties, log4j-node2.xml, and runNode2.cmd. For a three server setup, these are also needed: jppf-node3.properties, log4j-node3.xml, and runNode3.cmd.

Approximate model run times?

The runtime depends on the scale of demographic and socioeconomic data, the number of physical cores, and the speed of the processors. For year 2050, in the ARC one server setup, it runs for around 26 and 32 hours, depending on the server used.

MODEL SPECIFICATIONS

What sort of changes have been made to the network and zonal structure as part of the ARC ABM? Micro-zones or parcel-based data?

The model is based on 6,031 traffic analysis zones including external stations (5,922 internal zones). We are not planning on doing micro-zones, at least for now.

What additional land use/demographic data attributes are required for the ARC ABM vs. what was previously required for the 4-step model?

Before running ABM, Population Synthesizer needs to be run to produce the synthetic population and households. The required data include the number of HHs by size and income, employment by NAICS category, persons by age by county, labor make and use amount from PECAS (land use model), and university enrollment by TAZ. In the 4-step TBM, the input data is limited to HHs by size/income, SIC employment, etc. The model inputs are described in detail in the User Guide and in the Training Session 2 materials.

The Population Synthesizer generates a unique record for each household and person in the 21-county model area. Note that ARC can provide inputs for future forecasting years.

Any changes to the assignment model, especially to simulate tolling of managed lanes, etc.? DTA?

No DTA now, but possible in the future. Not much changed in highway assignment. Transit networks and assignments are modeled in Voyager's Public Transport (PT).

For tolling, the ARC ABM estimates a toll and non-toll split in mode choice. It does not apply a separate toll diversion method. We have developed a separate process to estimate tolls in the HOT lanes. Basically, the process tries to find the same transport cost between general-purpose lanes and toll lanes along the same corridor by adjusting tolls.

How does the Population Synthesizer distribute the Census data?

The model user specifies the number of households (by size and income level) that are to be synthesized for each TAZ. The Population Synthesizer takes this information as an input and allocates households from the expanded PUMS sample to each TAZ accordingly.

Is the employment completely independent of the Population Synthesizer?

ABM ensures that the number of workers produced by PopSyn is consistent with the employment numbers through the worker generation by households at each household size/income category and by shadow pricing. PopSyn doesn't explicitly match workers based on the employment data but the user can make adjustments to the controls to ensure consistency.

What year of dollars are used for income and other costs in the ARC ABM?

2010 dollars.

Why are there negative incomes?

In the ARC ABM, household income is an outcome of the population synthesis. Population synthesis can be understood as an expansion and discretization of a sample of households. The ARC Population Synthesizer uses the PUMS ACS 2007-2011 5% household sample. The PUMS ACS datasets report household income as collected by the Census Bureau.

The definition of income in Census data is the net income of the household. If a household has negative income, it means that over that year it accrued more business or investment losses than its other income. The ABM treats households with negative income as having no income. The Missouri Census Data Center's All About Measures of Income in the Census has a good description of what is considered as

income by the Census. The minimum income observed in the ACS 2007-2011 dataset for the ARC 20-county region is -$18,000. Out of approximately 100,000 household records in this dataset, 52 have income less than zero.

How are the income assigned to each household?

The income distribution of the synthetic population is determined by the income controls specified by the user. Note that the income distribution is controlled only at the level of income categories, as mentioned in the Activity-Based Model Specification Report. It is up to the user to specify income controls that reflect appropriate assumptions about future economic activity in the region. Further, note that there is no mechanism for specifying minimum or maximum income values in the synthetic population. The synthesis may or may not include the minimum and maximum values present in the seed household sample.

How will the ABM improve the accuracy of traffic forecasts over the 4-step trip-based model?

There is no guarantee that the ABM produces more accurate forecasts than the 4-step TBM in all cases. In terms of base year calibration, both models were well calibrated and validated. However, the ABM calibration took advantage of more recent data such as the latest household survey, the transit on-board survey, and NPMRDS (HERE) speed information that was not available when the TBM was developed. The ABM works at a more disaggregate level allowing for the use of more household/person characteristics that affect travel behavior.

Where to find the metadata for the activity-based dataset user to build the ARC ABM?

The data used in the estimation, calibration and validation of the ABM include the most recent regional household travel and transit on-board surveys, Census data, traffic counts, NPMRDS speed information, and transit ridership as reported by the transit agencies, among other sources. The employment estimates are prepared by ARC's Research & Analytics Division. The ABM also uses estimates of labor produced by the ARC PECAS land use model. These data sources are described in the model calibration and validation report.

Is there any documentation that provides an overview of how this model performs relative to ARC's previous modeling efforts?

No. Both models were well calibrated against data that existed at the time of the calibration of each model. Also, making comparisons to base year doesn't necessarily mean that one model is better than another. The most meaningful way to test the models is to do so by comparing how reasonable they are in forecasting mode.

This is done in one of two ways:

- Run both models for projects over a period of time and compare the forecasts to what happens as the projects open.
- "Back-cast" by using older data / networks and to "forecast" projects that have already opened

What software packages were used to build the ARC ABM? In particular, which statistics programs and versions were used?

The model estimation was performed using ALOGIT version 4. Various statistical packages were used to process the survey data and model outputs, including FoxPro, R, and SQL. Other customized data processing was performed in Cube, Python and Java.

Does the process of using random seed in many choices in the ABM affect the model results? For example, when running the same model multiple times, does each model produce different results?

When the model is run with the exact same inputs, it is designed to produce the same answer. If the random seed is changed, however, then the results will change even if no other inputs have been changed. Differences stemming from changes in random seeds are not significant – at most they are equal in magnitude to day-today differences in travel. These differences tend to be negligible at high levels of aggregation, such as freeway volumes, or county-level estimates.

Does the model inherently produce different results with successive model runs?

Indeed, our ABM implementation (much like any other ABM in the US and throughout the world) uses a random number seed in the Monte Carlo selection and stochastic simulation, hence making it nearly impossible to exactly replicate and recreate model results, so that's the nature of all ABMs, including ours.

For more info on the output variability caused by random seeds in an ABM, see:

- Output Variability Caused by Random Seeds in a Multi-Agent Transport Simulation Model
- Network Equilibrium with Activity-Based Microsimulation Models
- Modelling Public Transport Accessibility with Monte Carlo Stochastic Simulations
- Can We Control the Seed of a Monte Carlo Simulation to Recreate the Same Sequence of Results?

APPLICATIONS

What types of policy analysis and technical issues will the ARC ABM address, which cannot be addressed using the 4-step trip-based model?

ARC has used the ABM for the following scenarios that could not be addressed using the 4-step TBM:

- Telecommuting,
- Smart payment technologies (unified payment via smartphone),
- Increase in the use of transportation network companies,
- Aging population, and
- Autonomous vehicles.

ARC's current population forecast indicates that the metro Atlanta 20-county region is expected to add approximately 2.4 million people by 2040 when compared to the 2015 population. Was a land capacity analysis used for this projection?

ARC doesn't focus on general land capacity analyses because metro Atlanta is far from reaching the maximum development capacity in forecasting horizon. Therefore, the land capacity is not a major factor and constraint for population growth in the Atlanta region. Instead, ARC does use local zoning as the land capacity constraint in ARC's land-use modeling work. Local zoning may change over time, so ARC's forecasts reflect those changes. For example, ARC has developed the built-out scenario for Forsyth County under current zoning and development pattern.