Massachusetts Regional Household and Labor Force Projections and Subregional Allocation Documentation
Prepared for MassDOT Office of Transportation Planning

August 2023

Contents

INTRODUCTION

In May 2023, the Massachusetts Department of Transportation (MassDOT), the University of Massachusetts Donohue Institute (UMDI), and the Metropolitan Area Planning Council (MAPC) completed new population, household, and employment projections out to the year 2050 to support regional transportation modelling, project planning, and design. Population and industry projections at the regional level were created by UMDI, while household and labor force projections were developed by MAPC under contract to MassDOT. MAPC then allocated households and jobs to census blocks across the state using a land use model called UrbanSim. The resulting products paint a highly detailed picture of demographic change, industry shifts, and land development over the coming thirty years if current trends continue.

This document details MAPC's methods for the development of regional household and labor force projections, and for the allocation of households and jobs to sub-municipal geographies. UMDI's methods for the creation of regional population and industry forecasts are described elsewhere in documentation prepared by UMDI.

A simplified version of MAPC's projections process is as follows:

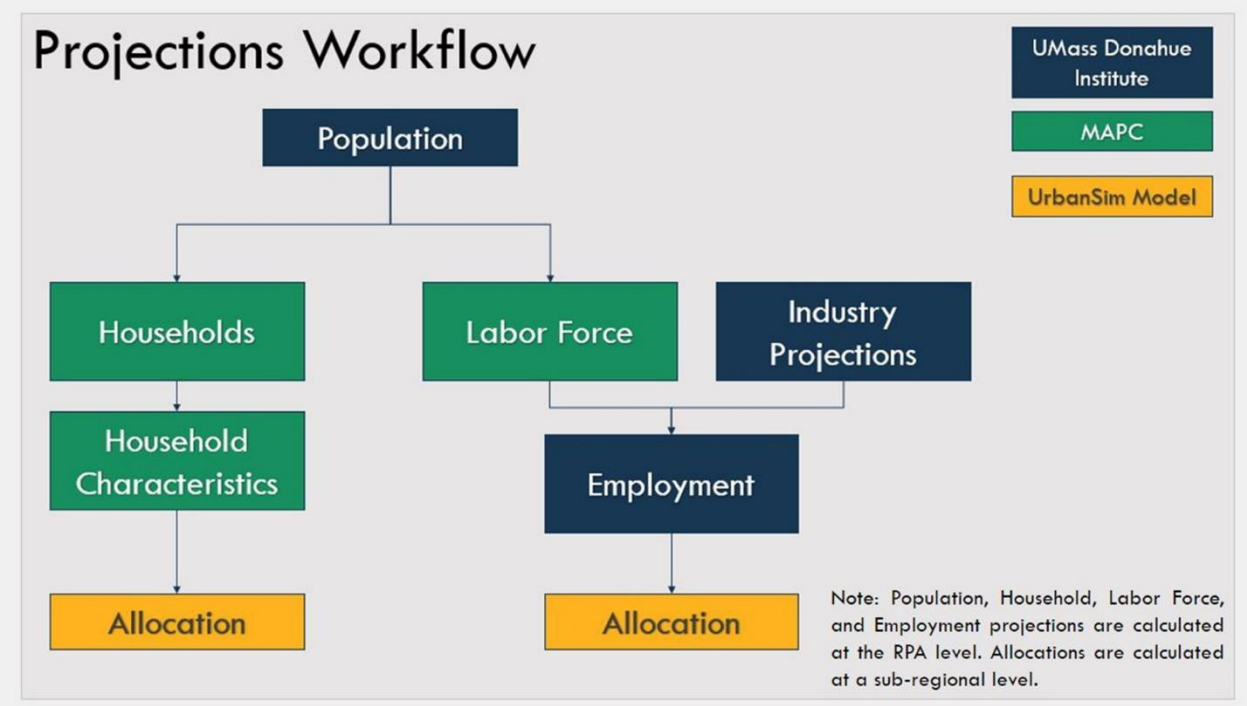
- Based on UMDI's regional population projections, MAPC first estimated the number of people in households by age group, and the number of people in the labor force for each Metropolitan Planning Organization (MPO) region and forecast year using values computed using data from the U.S. Census Bureau's American Communities Survey (ACS) 2015-2019 Public Use Microdata Sample (ACS PUMS). These estimates assume, generally speaking, that the likelihood someone of given age is in group quarters, in the labor force, or head of a household remains relatively steady over time.
- Labor force projections were further cross classified by age, educational attainment, and the labor force status of individuals. Household projections were further cross classified by age of householder, household type, number of persons in a household, and the presence of children in a household.
- The initial household estimates are further adjusted to match data from the 2010 and 2020 Decennial Censuses as closely as possible. This includes global headship rate adjustments to match 2020 household counts, targeted adjustments to headship rates for specific age groups and types, and formulation of synthetic households for each municipality in 2010 so that the base year households match 2010 census counts on all key characteristics.
- Regional labor force estimates, by age and educational attainment, were provided to UMDI to serve as a constraint on employment growth in the state and its regions. Overall employment in Massachusetts therefore grows at the same rate as the available labor force implied by the population projections.
- Once regional household and employment totals are finalized, the UrbanSim model allocates employment and households to sub-regional geographies. UrbanSim is a comprehensive urban simulation model that forecasts future growth of households and jobs based on location choice models that incorporate information about existing and planned development, access to amenities and labor, zoning and environmental constraints, and resident demographics. The model includes a "population synthesizer' that takes the aggregate household estimates prepared by MAPC and breaks them down into individual households with specific characteristics for size, workers, income, and children. Unlike prior models created by MAPC, growth is not allocated to municipalities and then to sub-municipal geographies; households and jobs are allocated from the regional control total directly to census blocks. As a result, there are

no fixed municipal totals—the forecast for any city or town may change as inputs to the model are modified.

- Following completion of the UrbanSim allocation, household population counts are reviewed and adjusted to ensure consistency with the regional UMDI Population projections for each age group. This process entails adding or removing people from selected microsimulated households to meet population targets without changing essential household characteristics.
- After this post processing, the microsimulated households are grouped by Transportation Analysis Zone (TAZ) and shared with the Central Transportation Planning Staff for travel demand modeling; and are summarized and published by municipality for other purposes.

The products can support the work of other state agencies and local governments in the preparation of other regional planning applications given the detail of the household and labor force characteristics. Data created via this process can be modified through changes to the various methodological assumptions (e.g., headship rates, labor force participation rates, education levels, etc.) to test alternative projection scenarios.

The following diagram summarizes the workflow for production of the 2023 LRTP projections:



REGIONAL POPULATION, HOUSEHOLDS, AND EMPLOYMENT PROJECTIONS

UMDI Population Projections

As with the last two rounds of statewide socioeconomic projections, MassDOT contracted with UMDI to produce population projections by sex and five-year age groups (e.g., 20–24-year-old) for each decade, out to 2050. UMDI's population projections represent the total population in each sex and age group for each Metropolitan Planning Organization (MPO) geography. The projections are based on a cohort component methodology that uses data about births, deaths, and migration to forecast future demographic change. Put simply, the model assumes the rate at which people of each age group in each MPO region move in, move out, have kids, and pass away stays about the same as it was during the period from 2015 – 2020. Over this period annual births declined while deaths, even prior to the pandemic, were rising. Domestic migration into Massachusetts declined throughout the period while international immigration was rising until 2017 before falling precipitously afterwards. On net, the state saw only small growth in its natural

increase (births minus deaths) in the population and a decrease in net migration by the end of the period used to anchor the population projections. This method neither assumes anything about major changes regarding where people choose to live or when they have children nor does it account for economic and employment factors such as housing prices, available jobs, or remote work. While we know all these things have changed since the pandemic, we neither know exactly how nor whether those changes will endure.

Given these caveats, it is useful to think of the UMDI projections as a "trends extended" scenario of Massachusetts' demographic future; one that represents *where the state was headed prior to the pandemic*. It's not necessarily the *most likely* outcome, but it provides a plausible and documented baseline for thinking about the future. With this baseline, we can better estimate how Covid-19 (or any number of future disruptions) will have effects distinct from those driven by massive demographic forces such as the aging Baby Boomers and the relatively small size of Gen Z compared to Millennials.

Information about the UMDI's population projection methodology and the MPO level population projections can be found here. It should be noted that UMDI methods use a cohort component method applied at the municipal level, the results of which are aggregated to the 13 MPO regions. In other words, UMDI produces municipal population projections and then compiles them into the regional totals needed by MassDOT and MAPC for this process. While UMDI's *municipal* population projections were created through the course of the MassDOT-funded effort, they are based on recent patterns of migration, births, and deaths and do not account for anticipated development, zoning constraints, or the impact of amenities and access on location choice – so they do not match municipal population in household projections prepared by MAPC. The municipal projections published separately by UMDI forecast the population for each municipality if recent trends are unconstrained and unaffected by land use and development factors.

Regional Households and Labor Force

Following receipt of the regional population projections from UMDI, MAPC's next step is to estimate the number and type of households likely to form in each region, as well as the number of people in the labor force. Households are classified by the age of the householder (I.e., the head of household), number of people in the household, presence of children under age 18, and income group (all income data is adjusted to a 2013 USD base year). MAPC also estimates the number of civilians in the labor force in each MPO region. MAPC's work achieves these ends using the Census Bureau's American Community Survey (ACS) data to construct multipliers which are applied to the population projections produced by UMDI. The following sections provide a detailed account of how these multipliers are created, implemented, and adjusted to create the input file for the UrbanSim microsimulation model.

ACS PUMS Geographies

Household and labor force multipliers for each region are based on ACS Public Use Microdata Sample (PUMS) data. Each PUMS release covers a five-year period and is available at a geographic level called a Public Use Microdata Area (PUMA) which contains at least 100,000 residents. This is done to maintain the privacy of ACS respondents, since the data made public in PUMS contains specific information about individual persons and households. PUMAs are usually larger than municipalities and generally do not adhere to county or regional boundaries (with a few exceptions). The geographies of PUMAs also change every decade; PUMAs boundaries for 2000, 2010, and 2020 are not always aligned with each other. Therefore, MAPC has constructed crosswalks that assign each decade's PUMAs to MPO regions. PUMAs for each region are then pooled to create a regional group of PUMS records used for household and labor force estimation. Please refer to Appendix A for the PUMA 2000 and 2010 to MPO crosswalks.

Household Population

To produce the projected number of households and people in the labor force, MAPC first converts the projected total population, created by UMDI, into the total population of civilians in households, excluding

military personnel and residents living in group quarters (dorms, correctional facilities, nursing homes, etc.) We adhere to the definition of "household" used by the Census Bureau (more information here) which is included as a variable in the PUMS data (see the PUMS data dictionary). To convert the UMDI data, we first determine the proportion of the population living in households as a share of total population by age group and MPO. Using data from the 2010 Decennial Census SF1 file, we subtract the total number of people (by five-year age group) in the military (since they do not count in the civilian population or live in households while on active duty) and in "group quarters" (more information here) from the total population and then divide the number of people in households by the total population by age and MPO. The resulting rate is the proportion of the population in households by age and MPO in 2010. To produce the **projected** household population, first, the total number of people in the military (using the 2010 data) are subtracted from the population in the forecast years. Then, the household population proportions are applied by age and MPO to produce the projected population in households for each projection year.

Labor Force

After estimating the projected household population by age group, MAPC applies educational attainment and labor force participation rates to each age group. Following consultation with MassDOT and the Massachusetts MPOs, MAPC used 2015-2019 ACS PUMS data to develop rates for the labor force model. This is the most recent pre-pandemic data available and represents what many economists consider a point in time the nation was at "full employment" (I.e., where virtually all residents willing and able to work are either employed or in between jobs). During this period, the statewide labor force participation rate was equal to levels that existed before the Great Recession and unemployment was at a 20-year low. Like the Bureau of Labor Statistics, MAPC uses these "full employment" conditions as an assumption for future labor force projections.

One factor in the labor force growth since 2010 was increasing participation rates for older individuals. The Bureau of Labor Statistics published a research report which concludes that the labor force participation rate (LFPR) for older adults is likely to continue increasing between 2020 and 2030. In consultation with Northeastern University Professor and UMDI partner Alan Clayton-Matthews, MAPC decided to adjust the LFPR for members of older age groups to reflect projected structural changes to the labor force. The adjustment factors for those are located here. It should be recognized that changes in labor demand, retirement fund volatility, and other factors may lead to other outcomes for the labor force participation rates of older adults. Furthermore, the projections do not make assumptions about cyclical factors that affect economic activity or further evolution of educational attainment or labor force participation rates for younger workers, both of which are highly uncertain given potential impacts of automation, artificial intelligence, climate change, international turmoil, and other factors. These topics can be explored in future economic scenarios.

The labor force model begins by estimating the likelihood (in 2015 – 2019) of a person in each age and sex category to achieve a given educational attainment level, classified as "High School or less", "Associate's degree or some college", "Bachelor's Degree", and "Master's or higher." These rates were applied to the future population by age and sex categories; then labor force participation rates for each age and educational attainment were applied to the number of persons at each education level. This rate is also derived from the 2015-2019 ACS PUMS data about "Employment status," for each age and educational attainment group, using records where the individual responded that they are either working for pay or seeking/available for work, and incorporating the older worker rate adjustments described above. Educational attainment and participation rates are constructed for each MPO region and applied to the household population data by age group and sex. The resulting output is the number of people in households in the labor force by sex, age group, and educational attainment.

It's important to note that the labor force participation rate derived from ACS PUMS data is not the same as the one derived from the Current Population Survey (CPS) which is generally used as the "official" rate

published by the Bureau of Labor Statistics (BLS). The PUMS-based definition diverges from the CPS definition in one key regard: The ACS instrument asks people if they are actively looking for work and available to take a job if offered one, but does not ask about the nature of the job search. The CPS questionnaire probes to see if people are actively looking for work (interviewing, calling contacts, etc.) versus passively looking for work (for example, looking at want ads in the newspaper). In the CPS, a person is unemployed only if that person responded affirmatively to having engaged in one or more active methods of job search. As a result, ACS estimates a larger number of unemployed respondents in the population—and a correspondingly higher labor force participation rate—than the CPS.

Household Types

For these projections, MAPC defines three general Household Types: a person living alone, households with two or more adults and no children, and all households with one or more children under 18. These types were chosen because they are assumed to have very distinct housing needs and household location choice preferences. Following consultation with MassDOT and the Massachusetts MPOs, MAPC used 2015-2019 PUMS data to develop region- and age-specific headship rates, defined as the probability of a person of a given age being the head of a given type of household. Headship rates for the three groups of households described above are applied to the total population in households for each forecast year. The resulting product is an initial estimate of households by age of head of householder and household type for each region and forecast year (2020, 2030, 2040, 2050). These household forecasts are used as targets in the PUMS "reweighter".

PUMS Reweighting

After determining the number of households and persons in the labor force, additional work is needed to create detailed household estimates that are needed for UrbanSim inputs. This detail is created using a process called "PUMS reweighting." The 'reweighter' is a method created by Alan Clayton-Matthews, Associate Professor of Economics and Public Policy at Northeastern University and Senior Research Associate at the Dukakis Center. This technique adjusts the household weights of PUMS data so that the resulting collection of weighted households matches a set of user-defined targets. The user can specify known characteristics of future residents/households and use the method to determine what more detailed characteristics of a synthetic future population would be. This method ensures consistency with projections created in different parts of the model. For a more detailed explanation of how the PUMS reweighter works see the UMDI employment projections documentation. The 'reweighter' has three input tables which act as targets for the future reweighted sample to hit: persons by age group; persons by educational attainment, age group, and labor force status; and households by householder age and household type. Age categories for the input tables are classified as 5-year age groups. The household types match the three household types created in the household model. The microdata sample to be reweighted are the 2015-2019 ACS PUMS data, grouped by the PUMAs that make up each MPO (see Appendix A).

Household Control Totals

The outputs of the reweighter are a list of PUMS "serial numbers" (unique identifiers for households), "sporder" designations (a number indicating the person record within a household), and the adjusted person and household weights for each forecast year. Those are merged to the original 2015-2019 ACS PUMS records. We then create a cross tabulation of the total number of households by the number of people in the household, presence of children in the household, the age of the head of household, and the household income group of the household for each year and MPO region. To interpolate data for the interdecadal years we use the linear interpolation method. UrbanSim uses this cross tabulation as the input to its household allocation model. The household control totals for the Statewide UrbanSim model have Duxbury, Hanover, Pembroke, and Stoughton assigned to the OCPC MPO/RPA whereas in the MAPC UrbanSim model they are included in the MAPC MPO/RPA. Additional adjustments to these estimates are described below.

The first adjustment harmonizes MAPC's household estimates with counts from the 2010 and 2020 Decennial Censuses. For a variety of reasons (geographic misalignment, changing headship rates, sampling errors, etc.), household estimates produced with PUMS data don't exactly match enumerated households in 2010 or 2020, even when using the same base population. Therefore, we compare the total number of households enumerated in the decennial census to the number estimated by our method. The ratio between these two figures becomes an adjustment factor applied to every age- and type- headship rate for those decades. As a result, MAPC's total household estimates match census data for 2010 and 2020 at the regional level.

For purposes of land use allocation (described below), MAPC interpolates the decadal household estimates/forecasts to develop an annual estimate of household counts for each household type for each year in the forecast period. For the post-2020 period, a linear interpolation between decadal household forecasts is adequate for modeling purposes. For the period from 2010 – 2020, adjustments were made to align household growth with observed housing unit construction activity. Household formation was slower in the earlier part of the decade and then accelerated as the pace of housing unit production increased to meet demand. In order to match estimated household growth to housing delivery, MAPC used the Massbuilds development pipeline database to estimate the cumulative total of new units built in each region in each year between 2010 and 2020. We calculated the proportion of housing units developed in or before each year compared to the total number of housing units in Massbuilds for the whole decade. The resulting yearly rates are applied to the total household change between 2010 and 2020, so that household growth matches the rate of housing production in Massbuilds. While Massbuilds is only a sample of housing units and a proxy for household formation, this method ensures that estimated household growth does not rapidly outpace observed housing production during the early part of the decade. Adjusted interpolation rates for each region can be found here.

Employment Control Totals

Employment forecasts for each region were prepared by the UMDI Population Estimates team and Alan Clayton-Matthews. Forecasts were prepared for each decade and for detailed industrial categories, then summarized to ten large "Super Sectors." To infer interdecadal data in years after 2020, we linearly interpolate the data between each decade and each Super Sector. For interdecadal years between 2010 and 2020, the data produced by the process described below is used instead of a linear interpolation. The ten Super Sectors are listed below and the constituent 2-Digit NAICS codes for each are listed in Appendix B.

SuperSector	NAICS Codes	Constituent Sectors	UrbanSim Model Segment
Construction	23	Construction	1
Education and Health Services	61, 62	Educational Services, Health Care and Social Assistance	2
Finance	52, 53	Finance and Insurance, Real Estate Rental and Leasing	3
Public Administration	92	Public Administration	4
Information	51	Information	5
Retail, Leisure, and Hospitality	44, 45, 71, 72	Retail Trade; Arts, Entertainment, and Recreation; Accommodation and Food Services	6
Manufacturing	31, 32, 33	Manufacturing	7
Other Services	81, 11, 21	Other Services, agriculture, mining	8
Professional and Technical Services	54, 55, 56	Professional, Scientific and Technical Services; Management of Companies and Enterprises;	9

		Administrative and Support and Waste Management and Remediation Services	
Trade, Transportation, Warehousing	22, 42, 48, 49	Utilities, Wholesale Trade, Transportation and Warehousing	10

The launch year for the employment projections is 2020; specifically, employment estimates for each region at the beginning of 2020, before any Covid-related job losses. These employment estimates were based on municipal-level employment (by sector) from ES-202. However, ES-202 data suppresses employment estimates for sectors with a small number of establishments in any given geography. As a result, the sum of the sector-level employment for a city or town may not match up to the total employment for that municipality; and the sum of sector-level employment for all cities and towns don't match up to statewide totals for that sector. To ensure that existing jobs in a given municipality or region were accounted for in the base year control totals, MAPC adjusted ES-202 employment estimates as follows: we used establishment-level data from Data Axle to augment sectoral municipal employment estimates where suppression was suspected; we compared municipal, county, and state data for each sector to impute how much employment was suppressed at each level; and then we distributed the imputed employment to municipalities based on the augmented sectoral totals. The result is a dataset in which municipal employment in each sector adds up to the state total employment. These municipal totals were then summed for each MPO region and provided to UMDI as the launch year employment for their forecasts.

SUBREGIONAL ALLOCATION AND SMALL-AREA FORECASTS

About UrbanSim

To develop small-area forecasts of population, employment, and land use change, MAPC has developed a customized implementation of the UrbanSim land use allocation model. UrbanSim is an established and validated land use model specifically designed to explore how changing demographics, economic growth, land use policies, and infrastructure investments may affect the development and character of cities and regions. As a modeling framework, UrbanSim has been adopted by nearly two-dozen cities and regions in the United States for land use forecasting. UrbanSim, Inc. makes the model available as a cloud-based subscription service called UrbanCanvas and provided support to MAPC for the estimation, calibration, and implementation of models customized to the Massachusetts and Metropolitan Boston context.

UrbanSim simulates the interactions among households, businesses, and developers within real estate markets. By modeling how households trade off housing costs, ease of access to jobs, and neighborhood amenities, UrbanSim simulates households' choices to select a housing location, housing type, and whether to rent or own. UrbanSim also simulates employment dynamics, including firms' location choices, as well as real estate developers' choices of what kind of buildings to build, where, and when, and whether to redevelop existing properties. Zoning and other policies constrain what developers can build, and assumptions about how transportation plans may affect the speed or cost of travel influence the attractiveness of different locations for households and firms. This influences prices, rents, and the market conditions for new development or redevelopment.

UrbanSim models individual households, jobs, and building construction, as well as their changes from one year to the next due to economic changes, policy interventions, and market interactions. UrbanSim simulates changes for each year, and the results of one year provide the starting point for the next simulation year. This method closely replicates the way that urban areas evolve, year over year, with mismatches between the supply and demand of housing and jobs.

Unlike prior land use models used by MAPC, growth is not allocated from regions to municipalities and then to sub-municipal geographies in a hierarchical manner. Rather, new households and jobs are allocated from

the regional control total directly to census blocks. As a result, there are no municipal totals that can be specified or maintained. When inputs to the model are changed (new development information, updated zoning, etc.), the entire regional allocation is re-run; even small, localized updates may have ripple effects that spread across municipal boundaries. In this way, the model simulates real-world conditions more accurately than an approach relying on fixed municipal 'control totals,', since major development projects and transportation improvements influence market and neighborhood conditions without regard to city and town boundaries.

Model Geography

MAPC developed two different UrbanSim models: one covers the 101 cities and towns in the MAPC region; the other encompasses the entire state. Location choice and price models for the two different model sets were specified and estimated separately; the MAPC model is tailored to the conditions at the core of the Metro Region, while the statewide model is better suited to forecast growth through the diverse regions of the state.

The statewide model is structured so that the user can provide regional control totals for each of the 13 Metropolitan Planning Organization (MPO) regions, and growth is allocated within each region. Four municipalities (Duxbury, Hanover, Pembroke, and Stoughton) are within the MAPC region but belong to the Old Colony Planning Council MPO. Regional totals, prepared at the MPO geography, can be input without alteration into the statewide model region for the OCPC MPO. For the MAPC model, additional steps are needed to account for the discrepancy between the 101 cities and towns in the RPA model region and the 97 cities and towns in the MPO control total region. To resolve this discrepancy, MAPC applies a proportional adjustment as follows. For the household controls, the MAPC model (MAPC101) cross-classified household data are multiplied by a factor that represents the proportional increase in the number of households in the MAPC101 geography relative to the MAPC97 geography. Factors differ in each year for which administrative data of the number of households in each region is available (2010-2020). Data in the forecast years have the 2020 factor applied. A similar track is taken for the employment control totals. ES-202 employment data is used to fill in the 2010-2019 baseline data for the MPO by Super Sector, since that data is available at the municipal-level. Using the 2020 ES-202 employment data, we construct an adjustment factor for the MPO level projections. As before, we calculated the proportion of employment in each Super Sector using the MAPC101 and MAPC97 aggregations. That proportion is then applied to the decadal forecasted data. For operational purposes, both models are run and the results of the statewide model for the Boston MPO's 97 municipalities are discarded and replaced with the results from the MAPC region model. (The full 101-municipality MAPC model results are used for regional analysis and planning not directly tied to the Long Range Transportation Plan.)

Within each region, the models allocate households and employment at the 2010 Census Block level, a total of nearly 55,900 blocks in the MAPC model region and 118,700 blocks in the statewide model. The model can output results at multiple summary geographies including counties, municipalities, 2010 Census Tracts, and Transportation Analysis Zones (TAZs) used for the CTPS Travel Demand Model. When summary geographies do not follow block lines completely (e.g., TAZs, some municipalities) an area-weighted division of the census block is used to divide the results.

The MAPC region model can accommodate specified sub-regional control totals for each of the MAPC Community Types. This allows MAPC to create scenarios with specified allocation of growth across the Community Types as opposed to one control at the regional level.

Model Inputs

Existing Households and Housing Units

The launch year of the model is 2010. It is essential, therefore, to have an accurate picture of housing units and households in existence at that time. Information about households and housing units was drawn from the 2010 Census and the 2009 – 2013 American Community Survey. MAPC used the reweighter described above to provide municipal level control totals by age of householder, household size, income, and household type to the UrbanSim model so that the household characteristics in the launch year match records from the Decennial Census. Launch year housing unit estimates for each block set to match the 2010 Census.

Employment

2010 Employment data by sector is drawn from record-level establishment data provided by InfoGroup (now Data Axle) in 2011. This dataset was reviewed and edited by MAPC and CTPS shortly after receipt in 2011 to resolve major discrepancies in employment and location. However, it was not reconciled with 2010 ES-202 employment data, so job counts do not match at the municipal, regional, or sector level.

MAPC also acquired 2019 employment data from Data Axle and performed extensive Q&A on this data to improve its accuracy. Data was provided to all MPO partners and extensive review and feedback was provided by MPO and municipal staff. The main applications of this information was to provide sub-municipal estimates of employment by sector for model calibration, and to fill gaps in the suppressed municipal ES-202 employment data. Data Axle data was used to help reconcile municipal ES-202 data with county and state estimates, so that the employment counts are aligned at all geographies and across all sectors. The 2019 employment estimates are used as the pre-pandemic employment counts for each municipality; in other words, the 2020 employment is effectively employment as of January 1, 2020, prior to the pandemic.

Block-level Capacity (Density and Developable Land)

A key assumption in the UrbanSim model is the number of housing units and jobs that could be accommodated within each census block. This assumption is termed "capacity" and is a function of two inputs: density assumptions (as dwelling units per acre or floor area ratio) and area of developable land. Multplying the density times the developable land area yields units and commercial area capacity. MAPC's Zoning Atlas serves as the principal source of density assumptions in the MAPC model. This database contains information for each of the 101 MAPC municipalities about maximum dwelling units/ acre and maximum commercial floor area ratio (FAR). Allowable density estimates were based on interpretation of local zoning codes/ordinances with a multi-step process to identify the highest allowable density if there are multiple development options. Detailed description of the interpretation methods can be found in the Zoning Atlas technical appendix. In January 2023, MAPC solicited input from all cities and towns in the region, asking municipal staff and boards to review the density estimates in the Zoning Atlas and provide feedback or updates as necessary. This feedback was incorporated into the assumptions used for the LRTP projections. It should be noted that not all overlay zoning districts were interpreted and incorporated in time for the LRTP projections and, therefore, the assumptions may over- or under-estimate allowable density in blocks where overlay districts provide more flexibility or impose additional restrictions stronger than the underlying base zoning.

To estimate developable land area in each block, MAPC removed areas with permanent and inviolable constraints that prevent any future development. These areas include the following:

- Permanently protected open space
- Water bodies and wetlands (source)
- Cemeteries

- Public Rights of Way

The presence of existing development was not considered to be a permanent constraint on development, since reuse and densification of most land uses is feasible if market conditions warrant. It should also be noted that the model estimates total capacity and compares this to existing development to determine the net additional units or commercial floor area that could be accommodated. If the existing development already meets or exceeds total capacity, no additional units can be placed by the model. An extensive review process was conducted by all cities and towns in MAPC to confirm the accuracy of these capacity constraints.

For the statewide model (SWM), MAPC collaborated with the state's other RPAs to compile a database of zoning district boundaries for non-MAPC municipalities. While some cities and towns provided current zoning data, others were unresponsive and MAPC resorted to using data originally published by MassGIS in 1999. Given the available timeline and budget, it was not feasible to manually interpret zoning bylaws and ordinances from the 250 non-MAPC municipalities. Therefore, MAPC and UrbanSim staff developed methods to estimate residential and nonresidential development densities based on recent development as recorded in MassBuilds and municipal assessor data.

We first looked at the existing residential densities in each zoning district using the MassGIS Level 3 Assessor Parcel Database. We then considered all recent developments over the last decade in each zone (MassBuilds developments as well any parcels in each zoning district where the 'built_date' was 2010 or later). Developments that fell under 40B permitting were excluded from this analysis. Also excluded were any outliers with densities or FAR greater than the 95% percentile of the initial estimates. These initial estimates based on "currently built densities" were then provided to all RPAs for detailed review. After extensive feedback and review from municipal and RPA staff, many additional direct manual adjustments were made to revise the initial estimates. The final spatial data on zoning district boundaries and assumptions regarding allowable densities used for the LRTP can be found here: Residential dwelling units per acre (maxDUA). Nonresidential Floor to Area Ratio (maxFAR).

For nonresidential development, the maximum FAR for commercial development in a zone was converted into a maximum employment capacity for each Census Block using the developed square foot-to-job factors described in more detail in the section below on the Massbuilds Development Pipeline. The final estimates for Census Block level capacities for residential units and employment used for the 2023 LRTP projections can be found here.

Travel Times and Job Accessibility

Estimated travel times and costs to different locations around the region are used as one input to the location choice models described below. This information comes from the CTPS TDM19 travel demand model. The outputs used for UrbanSim include travel times between each Transportation Analysis Zone pair for each mode and time of day. Travel times and costs are combined into a single measure expressed in dollar terms, with each hour of travel time equivalent to 18 dollars.

Once travel times are input, UrbanSim calculates the total number of jobs (of a given type) or households of a given type that are within a specified travel time/cost threshold. This value is recalculated for each model year/decade so that as new households and jobs are added, or as travel times increase or decrease, the accessibility to opportunities at a given location change over time.

Massbuilds Development Pipeline

While the UrbanSim model forecasts the location of future homes and job growth based on historic patterns, it can also incorporate information about development that is already planned or proposed ("scheduled development" in model terms.) Fortunately, MAPC already has a robust picture of recent and anticipated development, recorded in our collaborative online development inventory, MassBuilds. This database includes information about over 4,000 projects across the state either completed, underway, or planned. Information about each project includes residential unit count, commercial square footage (by various types), year complete, and point location, among other details. For the current forecasting effort, MAPC solicited input from all the municipalities in our region as well as the other twelve regional planning agencies in Massachusetts. Hundreds of records were created or updated through this process in early 2023. The Massbuilds export used for the LRTP projections can be found here.

To provide this data to the model, MAPC specifies the year the project is anticipated to be available for occupancy, the count of units, and the number of on-site permanent jobs it is likely to support. Jobs are estimated from commercial square footage using a range of factors (jobs per square foot of developed space) that depend on the community type, local job density, type of development (commercial, industrial, institutional), and employment sector, based on MAPC's analysis of pre-pandemic occupancy patterns and consultation with municipal and agency staff. At this time, UrbanSim accepts only total employment associated with scheduled development, not jobs by sector.

When added to the UrbanSim model, these developments increase the number of units or employment spaces available for occupancy, though whether they are filled with households or employees is determined by the household and employment location choice models. Scheduled development can also exceed the unit or job capacity for each block that was calculated from the zoning densities and developable land. In these cases, the model overrides the existing capacity and adds the development project as specified. Each project can be tagged with different scenarios, which allows us to create different model runs including/excluding certain projects based on assumptions about policy choices or likelihood of happening. While scheduled development adds units to a targeted block, there is no guarantee that those units will be occupied by households if other available units in the region are more attractive to the household location choice models. In a few select cases where it was judged that future development projects had a high likelihood of occurring and of being occupied, MAPC used "adjustments", in a procedure described below, to ensure that households were allocated to the relevant census blocks.

UrbanSim Structure and Sub-Models

A general overview of the UrbanSim block model (as well as details about data needs and specifications) can be found here: https://cloud.urbansim.com/docs/index.html

There are four main models in the UrbanSim system. Three of these are "location choice" models that drive the placement of new agents (households, jobs, development projects) in the simulation. Then there is a "price regression" model that estimates the variation of buildings' values. These models are further subdivided into submodels with different specifications depending on the category of the agent being modeled. For example, for the household location choice model (HLCM), households are grouped into 18 different 'segments' based on the age of the householder, the number of adults and/or children, and the household's income. Each segment's sub-model has a slightly different specification (predictor variables), and estimated coefficients for these variables.

Data sources for predictor variables include the Decennial Census, ACS, and PUMS data, Data Axle employment data, municipal assessor parcel data, and others. Measures of access to jobs and amenities are based on travel time and travel cost estimates (by mode) generated by the CTPS travel demand model, TDM19.

Household Location Choice Models

As described above, MAPC prepares cross-classified household projections for each decade year and interpolates the count of households with each unique combination of attributes for each year between the decadal regional totals. UrbanSim prepares synthesized households and population for the base year and then allocates new households for each one-year time step increment. When there is a net increase in the number of households in a given category, those households are sampled from the microsimulated household population and are placed into census blocks based on the HLCM. When there is a decline in the number of households in a given category, a corresponding number of simulated households meeting those criteria are selected randomly from across the region and are deleted from the model population.

Different household location choice models (HLCM) are specified for 18 distinct household segments defined by age of householder, household type, and income. (Initially 21 different segments were defined for modeling, but these were clustered for segments with similar characteristics and location choices.) The segment groupings are:

- Age of Householder: <35, 35-64, and 65+
- Household Income: <$35,000, $35,000-$100,000, and >$100,000
- Household Type: Single adult living alone, Multiple adults no children, One or Multiple Adults with children.

A full list of model specifications and coefficients is available here.

Each household segment has a unique combination of variables and coefficients in the location choice model used for estimating where new households will settle. The following variables are used in one or more of the HCLMs for the MAPC model region:

- Household density (2010 Census Block Group)
- Proportion of units in single family vs. multi-family buildings (2010 Census Block, MA Land Parcel Database)
- Mean rent (2008 – 2012 ACS Census Tract)
- Mean home value (2010 Census Block Group, Warren Group Real Estate Data)
- Mean year built, housing units (2010 Census Block Groups, MA Land Parcel Database)
- Total jobs within 45 minutes by transit, AM peak period (Traffic Analysis Zone (TAZ), CTPS Travel Demand Model)
- Total households within [8 dollars in time terms] by auto (TAZ, CTPS Travel Demand Model)
- Total jobs within [7 dollars in time terms] by auto (TAZ, CTPS Travel Demand Model)
- Mean income (2008 – 2012 ACS Census Block Group)
- Census Block is within ¼ mile of transit station (2010 Census Block, MassGIS MBTA stops)
- Census Block is within 3000m of a major road (MassDOT Road Inventory)
- Percent of roads with sidewalks (MassDOT Road Inventory)
- Number of high-income households within 1500 meters (ACS 2008 – 2012 Census Block Group)
- Count of units built since 2010 within 1500 meters (SF-1 2010 Census Block Group, ACS 2008 – 2012 Census Block Group)

Employment Location Choice Models

The ELCM is comprised of 10 different employment location choice models, one for each SuperSector described above. The model was estimated using 2011 establishment-level employment data from

InfoGroup (now DataAxle), which was reviewed and corrected by MAPC and CTPS for prior modeling efforts. As with households, the decadal control totals for employment are interpolated linearly for each sector to create projected employment counts for each individual forecast year. Similarly, the model places any net increase in a sector's employment into blocks using the appropriate ELCM, whereas declines in a given sector are resolved through random selection and deletion of jobs in that sector across the region.

- Residential unit density (2010 Census Block)
- Residential household density (2010 Census Block)
- Proportion of units in single family vs. multi-family buildings (2010 Census Block, MA Land Parcel Database)
- Mean year built, housing units (2010 Census Block Groups, MA Land Parcel Database)
- Total households within 8 dollars in time terms by auto (TAZ, CTPS Travel Demand Model)
- Total jobs within 8 or 25 dollars in time terms by transit, AM peak period (TAZ, CTPS Travel Demand Model)
- Census Block is within ¼ mile of transit station (2010 Census Block, MassGIS MBTA stops)

Model specifications and coefficients are available here.

Residential Development Project Location Choice Models

The MAPC Region UrbanSim model also relies on a pair of residential development project location choice models (RDPLCM). There is one model for Single Family development and another for Multi Family development. These models are used to estimate the likelihood that a hypothetical developer will build new units of either type in a given census block. The RDPLCM is run as the last sub-model in a given year, so the resulting housing units are available in the subsequent year for households to choose in the household location choice model. The following variables are used in either the single family or multifamily RDPLCM:

- Proportion of units in single family vs. multi-family buildings (2010 Census Block, MA Land Parcel Database)
- Household density (block groups, 2010 Census)
- Mean year built, housing units (2010 Census Block Groups, MA Land Parcel Database)
- Number of units built since 2010 within 800 meters (block groups, 2010 Census, 2013-2017 ACS)
- Census Block is within ¼ mile of transit station (2010 Census Block, MassGIS MBTA stops)
- Total jobs within 8 dollars in time terms by auto (TAZ, CTPS Travel Demand Model)
- Number of high-income households within 1500 meters (Census Block, ACS 2008 – 2012)
- Mean rent (Census Tract, 2008 – 2012 ACS)
- Ratio of households to residential units AKA vacancy rate
- Total households within $35 travel time via transit during AM peak (CTPS travel demand model)

Model specifications and coefficients are available here.

Real Estate Price Models

The MAPC Region UrbanSim model relies on a pair of real estate price models (REPMs) to estimate residential sales prices and rental rates. The value model was estimated using Warren Group transaction data and the Rent model was estimated using ACS data.

- Residential unit density (2010 Census Block Group)

- Mean household income (2008 – 2012 ACS Block Group)
- Proportion of units in single family vs. multi-family buildings (2010 Census Block, MA Land Parcel Database)
- Total households within 8 dollars in time terms by auto (TAZ, CTPS Travel Demand Model)
- Census Block is within ¼ mile of transit station (2010 Census Block, MassGIS MBTA stops)
- MAPC Community Type
- Mean year built, housing units (2010 Census Block Groups, MA Land Parcel Database)
- Census block is within 3000m of a major road (MassDOT Road Inventory)

Model specifications and coefficients are available here.

Model Calibration

The models above have all been calibrated using data on the change in households and employment between 2010-2020. The household location choice models are calibrated using the observed household growth estimates between Census 2010 SF1and Census 2020 PL-94 redistricting data, at the municipal level. In the context of UrbanSim, 'calibration' involves making small adjustments to model coefficients to minimize the value of an objective function (in the case of households, for example, the calibration seeks to minimize the difference between the model's predicted household growth from 2010 to 2020 with the household growth reported by the Census Bureau, for each municipality between 2010 and 2020). After calibration, the model estimates of total households by municipality in 2020 are within a few percent for nearly every municipality in the state.

The employment location choice models were calibrated using the combination of 2019 ES-202 data on employment by municipality and 2019 DataAxle business location and employment data described in the **Employment Control Total** section earlier in this document. The residential development project location choice model was also calibrated using ACS 2009-2013 data.

As stated above, UrbanSim's calibration methods do not involve adding an error term to achieve a perfect fit at every geography. Instead, coefficients of the estimated models are iteratively changed to achieve a better fit with the calibration targets. This means that the 2020 forecast does not exactly match 2020 observed households or jobs at the sub-municipal level. When using the projections, it is advised that users look at the change between 2020 and a forecast year, rather than using absolute numbers for 2020 or afterwards.

Adjustments

While households and job locations are primarily determined by the location choice model, a user can also specify a particular number of households or jobs of a given type be allocated to a specific block in a specific year. These are termed 'adjustments' and are specified for a specific start year and end year range. Adjustments can be specified on a single variable such as job sector, age of householder, or household income. Cross classified household adjustments (e.g. 100 senior-headed households with incomes <$50,000) are not possible at this time.

An adjustment applies to the total number of households and jobs in a given block, so existing households and employment must be taken into consideration when applying adjustments. For the year range provided, the model will hold the block households or employment at the specified levels in the adjustment. After the end year of the adjustment, the model is then free to add or remove households or jobs using the standard allocation models.

MAPC used adjustments for two purposes: to improve the post-calibration fit with municipal and sub-municipal households and jobs for 2020; and to ensure that households are allocated to large future developments that are not fully occupied by the HLCM. A full enumeration of all adjustments applied to the LRTP projections can be found here.

Vacancy Rate Targets and Unavailable Vacant Units

UrbanSim allows users to specify a target regional vacancy rate, which determines the number of units built by the RDPLCM. For purposes of the 2023 LRTP projections, we applied a regional vacancy rate of 4.5% in both the MAPC and statewide models. However, the model does not distinguish between available vacant units (those that are available for sale or for rent) and those that are vacant but effectively unavailable to home-seekers. This latter category includes seasonal or second homes, vacation rentals, units that have been sold or rented but not yet occupied, units used for commercial purposes, and housing for migrant workers. In its default state, the model will allocate households to these "unavailable vacant" units even though they are not actually available for occupancy, which then results in curtailed production of new units through the RDPLCM. Since there are so many vacant units that are, as far as the model is concerned, available for occupancy, it does not then build enough units to keep up with household growth.

To address this issue, MAPC undertook a process to distinguish available from unavailable vacancies at the municipal level and remove the latter from the inventory of existing units provided to UrbanSim. We used information from the 2010 Census and the 2015-2019 ACS to determine the share of vacant units in each municipality that were classified as seasonal or occasional units, sold or rented but not yet occupied, or other vacant units. We then applied this share to the total vacancy in each census block and removed that many units from the total housing stock. This estimate of unavailable vacant units is held constant at the block and municipal level over the forecast period and is added back to the projections of housing units to produce estimates of total housing units in a forecast year. This prevents the model from filling existing non-available vacant units with new households, resulting in vacancy rates for the 'available' unit counts that are more reflective of reality.

Post-Processing of Household Population

As described above, MAPC provides detailed household forecasts to the UrbanSim population synthesizer, which enumerates individual households matching the characteristics of the target totals. The provided households are cross-classified by age of householder, presence of children under 18, household size (1, 2, 3, and 4 or more people), and other characteristics. While household counts and characteristics are matched, the population contained within the synthetic households may exceed or fall short of the population in household forecasts for any given forecast year and age group. **UrbanSim does not utilize population control totals.** As a result, the microsimulated population in households that are allocated to match the **household control totals** will always differ somewhat from the population totals provided by UMDI.

Due to the nature of the reweighter and population synthesizer, it is not possible to configure these tools so they converge on a population that exactly matches the UMDI forecasts. The biggest discrepancies were underestimates of the population in the 15-34 and 65-74 year old age groups, though there were also over-and under-estimates for other age groups. As a result, MAPC took a multi-step and iterative process to align the synthetic household population with the forecasted population.

Household Size Adjustments

The first step entailed adjusting the distribution of household sizes provided as inputs to the population synthesized. Based on a review of the synthetic population age distribution and tests of the shift/delete/duplicate methods described below, we determined that adjustments to household sizes were needed to achieve convergence without resulting in unreasonably large household sizes. Specifically, we reduced the number of single-person households and increased by a corresponding amount the number of

households with multiple adults and no children. The effect is that the same number of households incorporate a larger population. This process was applied to two age ranges for the head of household: the 15-34 and 65-74 age groups.

- For the 15-34 age group the shift is implemented as such:
    - For households with three persons and no children in the household, the number of households across all income group segments was multiplied by 1.3 (a 30% increase).
    - For all households with four or more people and no children in the household, the number of households across all income group segments was multiplied by 1.5 (a 50% increase).
- For all single households, the number of households was decreased by the sum of the increase in three- and four or more-person households. The decrease is distributed proportionally across all income group segments based on the pre-adjustment income group distribution of households.

For the 65-74 age group the shift is implemented as such:

- For households with two-, three-, and four or more-person households with no children, the number of households across all income group segments was multiplied by (1+([Headship rate for households with two or more persons without children])
- For all single households, the number of households was decreased by the sum of the increase in two-, three-, and four or more-person households. The decrease is distributed proportionally across all income group segments based on the pre-adjustment income group distribution of households.

The next adjustment is applied to the microsimulated population after allocation. The regional control total household population and the UrbanSim microsimulated population are compared to determine discrepancies for each five-year age group, for each MPO, and each decadal forecast year.

Age Shift Adjustments

In regions and forecast years where there are adjacent age groups with discrepancies different in sign (one microsimulated age group has a surplus compared to the regional control totals and an adjacent age group has a deficit of population), then population is shifted to the adjacent age group. In this process, the age variable for individual person records from an age group with a surplus of population are altered by plus or minus five years to shift them into an adjacent age category. For example, if the 25-29 age group has a surplus population and the 30-34 has a deficit, enough 25-29 year old records are randomly selected and 5 years are added to their age to eliminate either the surplus or the deficit. After this process is run, MPO age group segments with remaining surpluses or deficits are passed to the "Duplication/Deletion" script.

Person Record Duplication/Deletion

In the next step of the process, MPO age group segments with outstanding surpluses and deficits are brought into alignment with the UMDI household population data by randomly selecting household members for duplication or deletion from a subset of persons who represent the characteristics in the delinquent MPO age group segments. Before the duplication/deletion process occurs, a subset of household members in each segment is created. Depending on the age group, the criteria for inclusion in the subsets differ. For example, the duplication subset for children is constrained to households with fewer than four previously existing children (so as not to create households with an inordinately large number of children) and the deletion subset for children is constrained to households with more than one child so as not to alter the household type of said household (i.e. to go from a household with children to a single person household or a multiple adult household without children). Duplication/deletion subsets for adults in households are less constrained. The addition subset for each MPO age group is dictated by household type (multiple adult households without children) and age of the household reference person (persons can be added to households with a reference

person who is 12 years younger or older than them). The deletion subset is constrained by the household type and the number of preexisting persons in the household. Then the microsimulated data for each MPO age group segment has person records either duplicated or deleted until parity with the UMDI household population data is reached. The result is microsimulated data that reflects the original MPO-level household population targets used in development of the household and labor force data used at the outset of the projections process.

Group Quarters Population

As described previously, group quarters residents are subtracted from the total UMDI projections prior to estimating and allocating future households. In order to estimate total population at the municipal level in future years, it is necessary to allocate forecasted group quarters population from the region to municipalities. The method for this allocation is quite simple: MAPC calculates the share of the MPO group quarters population, by age group, residing in each municipality in the year 2010 (using decennial census data.) This share is applied to the future year forecasts of MPO-level group quarters population by age; the result is an estimate of group quarters residents by age in each municipality in each forecast year, which sum to the MPO-level group quarters forecasts by age.