Chicago Metropolitan Agency for Planning

ON TO 2050
plan update

Socioeconomic forecast
appendix

Contents

Introduction **3**
Part 1: 2050 regional socioeconomic forecast update **4**
Introduction 4
Demographic model: Data, methods, and assumptions 5
Economic forecast: Data, methods, and assumptions 11
Integration of demographic and economic models 12
Socioeconomic forecast: Regional results 14
Part 2: 2050 Sub-regional forecast update **18**
Introduction 18
The UrbanSim land use model 18
Model data requirements 20
Policy influence and implementation 26
Socioeconomic forecast: Sub-regional results 30
Appendix 1: UrbanSim developer model overview **35**
Appendix 2: UrbanSim estimation/calibration **38**
Appendix 3: Building types used in UrbanSim **43**
Appendix 4: Adjustment areas **45**

Introduction

This document provides an overview of CMAP's socioeconomic forecasting process and results in support of the ON TO 2050 plan update. High-level results will be presented here. A more complete set of data tables will be made available on the CMAP Data Hub at: https://datahub.cmap.illinois.gov/dataset/2050-forecast-of-population-households-and-employment after plan adoption in October 2022.

Socioeconomic forecasts are a required element in a metropolitan planning organization's long-range transportation plan, with a horizon year that is at least 20 years out from the plan's adoption date.[1] The results serve dual purposes: they provide an understanding of forecasted population and employment trends to help shape plan recommendations, and they serve as an input to CMAP travel models for air quality conformity analyses, as well as for small-area traffic projections.

The forecast has two major components: the regional socioeconomic forecast and local projections, which are the disaggregation of regional totals down to the local level. These two exercises draw on different disciplines. The regional forecast is an exercise in demographics and macroeconomics, while the local forecast is a spatially oriented exercise that requires more of a focus on local constraints to growth, transportation accessibility, real estate supply, and a host of other factors. The next two sections describe these processes in greater detail.

While the forecast is driven by transportation planning needs, these projections are also used by CMAP staff, as well as by partner agencies, local communities, economic development organizations, and watershed planners. In acknowledgement of these diverse needs, CMAP is committed to providing results with more demographic and temporal detail beyond basic travel model requirements. While much of this detail is limited to the regional totals, it does provide an overview of general demographic trends in northeastern Illinois forecasted for the coming decades.

Note: Like most forecasting efforts, many of the underlying assumptions behind this forecast are trend-based and do not account for unanticipated behavioral changes (let alone major disruptions, such as the COVID-19 pandemic). As such, the best one can claim for a forecast is that it is defensible and based on reasonable assumptions.

[1] U.S. Government Publishing Office, *Electronic Code of Federal Regulations*, Title 23/Chapter I/Subchapter E/Part 450.324 https://www.ecfr.gov.

Part 1: 2050 regional socioeconomic forecast update

Introduction

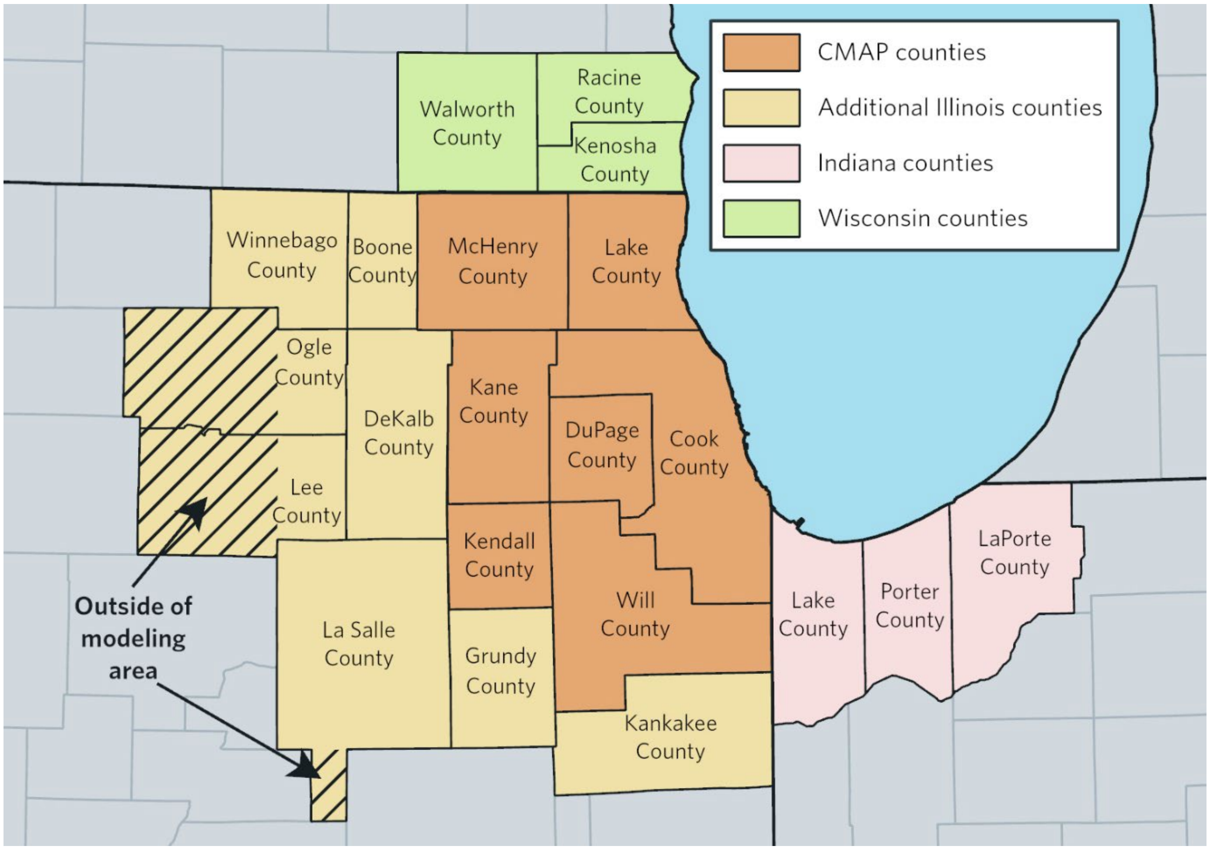
The technical approach for the regional socioeconomic forecast mirrors that of the original ON TO 2050 forecast, which was produced by Louis Berger, Inc. CMAP staff split the project into two contracts for the plan update cycle, the goal being to develop a sustainable process that will allow staff to generate forecasts in-house in subsequent plan cycles. To that end, CMAP entered into an inter-governmental agreement with the Applied Population Laboratory (APL) at the University of Wisconsin-Madison to review the Berger approach to the demographic forecast and advise on best practices and data sources so that CMAP could develop a stand-alone demographic model for continued use. CMAP also contracted with the firm EBP to provide the agency with an employment forecast based on projections from Moody's Analytics with modifications to account for differences in employment sector definitions and adjust for accounting of temporary workers, as well as to provide documentation and source code so that the agency can replicate these processes in the future.

Much of the data used to inform the population and employment models are derived from state- and county-level sources. As a regional planning agency, CMAP recognizes the importance of inter-county dependencies and that it is unreasonable to expect counties to grow in isolation of one another. As this is a regional forecasting exercise, all county-level inputs from the demographic and employment models were summed into regional totals for the regional forecast; sub-regional (county, township, travel model zone) output totals were generated through the local forecast process, described in Part 2 of this document.

Part 1 will discuss the methods, data, and assumptions behind the demographic and economic models, followed by the steps taken to reconcile the results of the two models. Regional population, household, and employment projections are presented at the end of this section.

Note: While the regional forecast was developed specifically for the seven-county CMAP region, this effort produced projections for a wider, 21-county area (**Figure 1**) coinciding with the area modeled by CMAP's travel demand models. The broader area is necessary so that the entire modeling area is informed by a forecast produced with consistent methodology. Results for areas outside of the seven-county CMAP region are used solely as travel model inputs and are not considered part of the official ON TO 2050 forecast; therefore, they will not be reported here.

Figure 1. Twenty-one county modeling area



Demographic model: Data, methods, and assumptions

CMAP's population projections for 2025–2050 are founded on an established demographic technique called the cohort component method. In short, this method analyzes the historical patterns of the primary elements of population change — fertility, mortality, and migration — and extends them into the future either by trending these past indicators or harnessing them to predictions at a larger geographic scale, such as a state or the nation. Recognizing that it is difficult to choose and justify the scale and timing of major trend changes in long-term forecasting, CMAP's projection methodology tended toward continuation of existing observed trends plus additional rate change modifications in order to avoid extreme, unsustainable, or illogical results within the forecast period (an example of this is the age 15–19 fertility rate in **Figure 2**).

In addition to the cohort component process, CMAP applies a labor force model to supplement net migration assumptions. This model incorporates employment projections to reconcile the working-age population (labor supply) with the anticipated labor force demand.

Data collection, formatting, and projection were performed in R, an open-source statistical programming language.[2] Census data were retrieved and analyzed using tidycensus and the tidyverse collection of R packages; charts were produced with ggplot2 (a graphics package for R), and cmapplot, a CMAP-developed customization of ggplot2 that applies agency graphics standards to R products.

The major elements of the demographic model are described below, including data sources and assumptions developed in consultation with APL. Additional detail is available in the report *Demographic forecast technical report, ON TO 2050 plan update*, available on the CMAP Data Hub after plan adoption in October 2022. Much of the text below is excerpted from this report.

Births/fertility

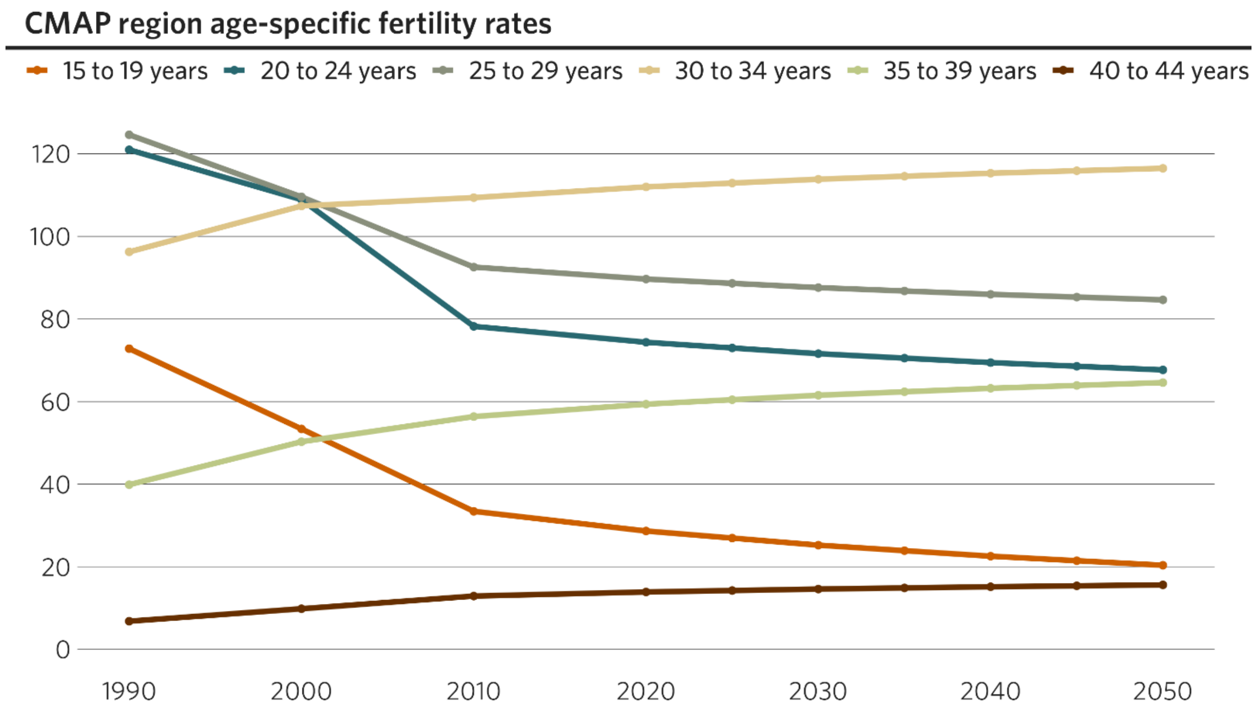
Data on births and deaths were obtained from the Illinois Department of Public Health, Indiana State Department of Health, and Wisconsin Department of Health Services to develop fertility rates, as well as mortality rates discussed in the next section. Data on births were used to calculate age-specific fertility rates (ASFRs) for the years 1990-2010, grouped into six age cohorts (in five-year ranges from ages 15-19 through 40-44). Historic rates showed two distinct trends: fertility rates for the under-30 cohorts showed a marked decrease, while all age-30-and-above cohorts showed moderate increases. These trends were carried forward into the forecast years using a logarithmic trend projection, which allows for a gradual slowing of trends in acknowledgement of the uncertainty of these trends carrying forward into the future. **Figure 2** (below) depicts the historic (1990-2010) and projected (2020-2050) fertility rates by age cohort, reported as live births per 1,000 females. Total births are generated by multiplying projected ASFRs by the projected number of women in each age range.

Deaths/mortality

The most common single-number metric for discussing mortality is life expectancy at birth, typically calculated separately for men and women because each sex faces different mortality risks across their lifetimes. Life expectancy is a synthesized one-number estimate based on the mortality rates (or, conversely, survival rates) of age-specific cohorts over a period of time such as one year, five years, or a decade; life expectancy aids our understanding of a geographic area's mortality patterns through time or in comparison among geographies (see **Figure 3**, below).

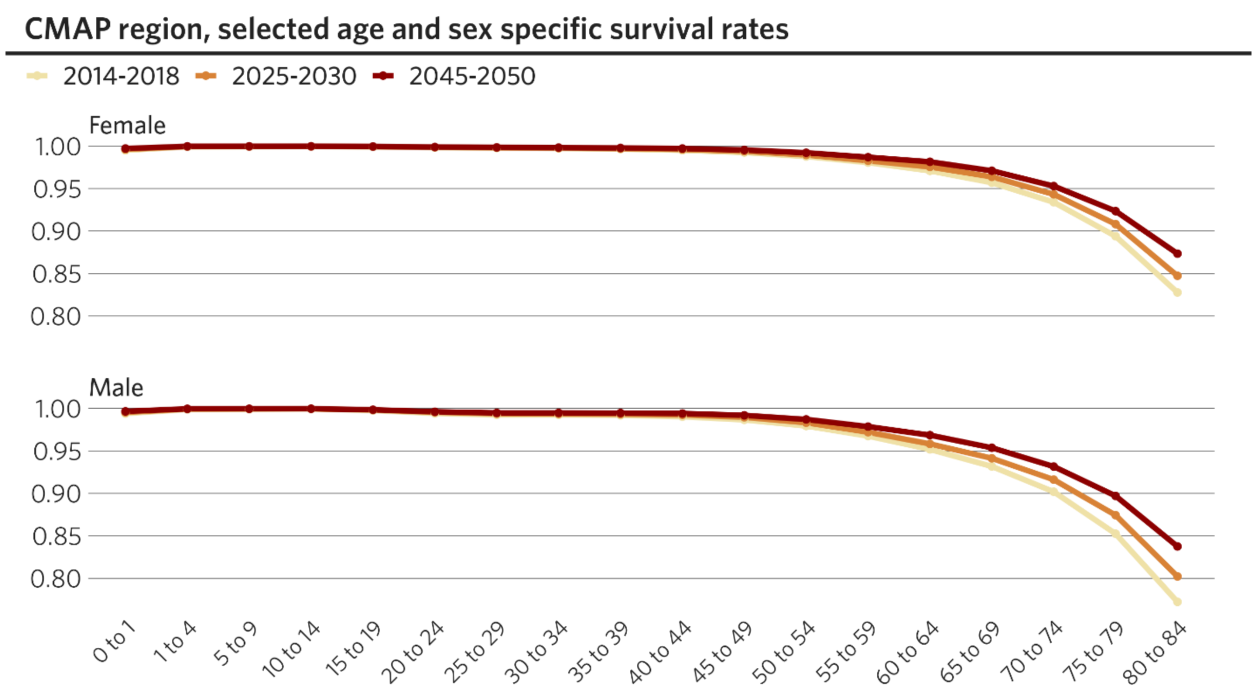
[2] R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Figure 2. Historic and projected fertility rates for the CMAP region



Source: Census Bureau 1990-2010, CMAP Demographic Projection 2020-2050

Figure 3. CMAP region, selected survival rates by age and sex



Note: 2014-2018 mortality rates are derived from local health department data. 2025-2030 and 2045-2050 rates are projections based on local data and Social Security Administration national projections.

From 1990 through 2010, life expectancy in the CMAP region showed strong gains, increasing 5.7 years for men and 3.6 years for women, while the gap between men and women shrank from 7.0 years to 4.9 years. Even from 2010 to the 2014-2018 period — the most recent period age- and sex-specific death data could be collected from county health departments — life expectancy improved for both men and women.

Across the forecast period, following national mortality patterns predicted by the Social Security Administration, average life expectancies in the CMAP region are expected to continue to increase, albeit more modestly than in recent decades.[3] Male life expectancy at birth may reach 80.7 years in the 2045-2050 period, while female life expectancy may reach 84.9 years. As with recent history, the projections indicate that the male-female life expectancy gap will continue to shrink, from 5.1 years in 2014-2018 to 4.2 years by 2050.

Migration

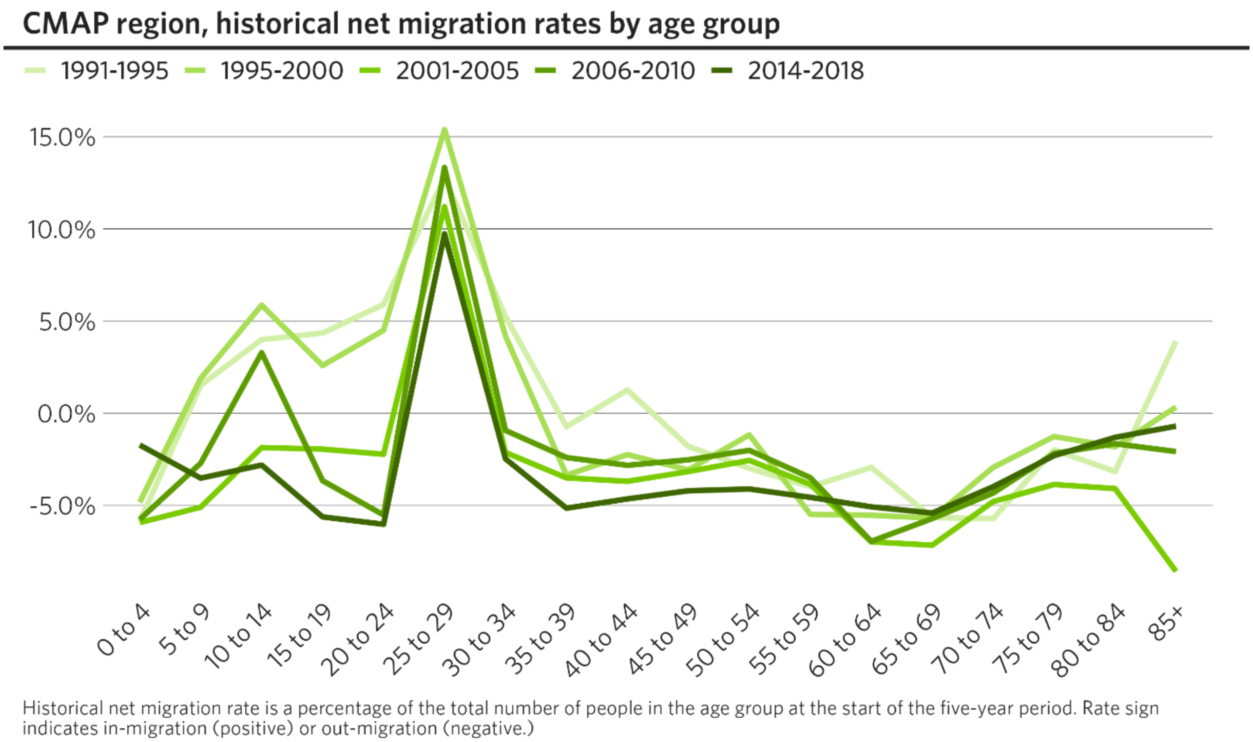
The measurement of migration is best understood as a "net" process — people migrate into an area over a period of time, and others move out of that same area; the net gain or loss due to migration is the result of the in-flow minus the out-flow. Unlike births and deaths, which are recorded as official vital events, measures of in- and out-migration are difficult to ascertain. Thus, for the purposes of population projections, net migration is calculated through a residual process: it is estimated as the difference between the total population change and natural increase (births minus deaths).

Migration itself consists of two streams: domestic (within the United States) and international. Like gross in- and out-migration, obtaining accurate records or estimates of these domestic and international components is difficult. For these projections, only total net migration values are calculated. The region's net migration, in total, has varied extensively across the past three decades. After a decade of positive net migration in the 1990s, the region experienced net out-migration during the 2000s and 2010s.

Net migration also follows particular patterns based on age and sex; the age- and sex-specific rates form the basis of projected migration. For the 2014-2018 base period, estimates of the CMAP region's net migration were calculated using births, deaths by age and sex, and the total estimated population by age and sex.

[3] Social Security Administration, "2019 Trustees Report." https://www.ssa.gov/oact/HistEst/PerLifeTables/2019/PerLifeTables2019.html

Figure 4. Historic net migration rates by age for the CMAP region



Because recent total net migration was negative (see 2014-2018 migration in the graph above), many of the migration rates by age and sex are negative. One distinctive difference is the consistently strong net in-migration for residents ages 25-29. Like most larger metropolitan areas in the country, the Chicago metropolitan region attracts many young post-college graduates. These figures form baseline net migration assumptions, which are then adjusted based on regional economic expectations (discussed in the demographic and economic integration section later in this chapter).

Additional demographic variables

The following variables were developed as additional model outputs to form a more complete demographic picture of the region.

Household formation

The calculation of households from population totals is an important component of the forecast, as the number of households is a key input to additional CMAP analyses and products. Due to delays with the 2020 census, data from the 2010 census were used to determine age-specific headship rates, which help account for changing household formation patterns and changes in the age distribution of the total population. Headship rates are applied to the household population to calculate the total number of households. The summed total of households by all cohorts yielded the total number of households for each forecast year.

Household totals are also summarized into three age-of-householder categories for travel model purposes.

Race and ethnicity

Estimations of the population by race and ethnicity are broadly summarized by five non-overlapping categories: Hispanics of any race, non-Hispanic white persons, non-Hispanic Black persons, non-Hispanic Asian persons, and non-Hispanic persons of other or a combination of races.

Like the ON TO 2050 forecast approach, a mixed log-linear rate projection approach was applied to age-specific race and ethnicity data collected in the 1990, 2000, and 2010 censuses. This approach accounts for the recent rapid growth of certain groups but moderates the rate of growth across the 30-year projection period to avoid illogical results by the year 2050, and balances recent observed trends with uncertainty about future changes in population change.

The Census Bureau has identified changes in survey design and self-identification as significant factors for recent trends in race/ethnicity estimations.[4] Further work is necessary to better characterize, understand, and project demographic trends in the CMAP region.

Group quarters populations

All demographic modeling described up to this point addressed only the household population. To develop a total population estimate, projections must also account for group quarters populations, both institutionalized (nursing homes, prisons, etc.) and non-institutionalized (college dormitories, military quarters, etc.). To estimate the change in group quarters populations, 2010 census data were used to calculate the proportions of people in group quarters arrangements relative to the 2010 total population. These proportions were applied forward with the exception of the military quarters population (located exclusively in North Chicago at Naval Station Great Lakes), which was held constant.

[4] U.S. Census Bureau, "2020 Census Illuminates Racial and Ethnic Composition of the Country." 12 Aug 2021, https://www.census.gov/library/stories/2021/08/improved-race-ethnicity-measures-reveal-united-states-population-much-more-multiracial.html

Economic forecast: Data, methods, and assumptions

In August 2020 CMAP issued RFP #243, "Regional Employment Forecast," and selected the firm EBP to develop the economic portion of the regional socioeconomic forecast. Scope requirements were:

- Report by NAICS two-digit sector using Bureau of Labor Statistics definitions
- Reallocate employment totals from NAICS sector 561320 (Temporary Help Services) into the sectors that temp workers actually work in; provide totals for reallocated and un-reallocated
- Report total employment (including self-employed), as well as wage and salary only
- Develop a baseline/likely scenario along with low/pessimistic and high/optimistic scenarios reflecting the uncertainty that typically surrounds employment forecasts

An overview of EBP's approach to the employment forecast is presented below. For a more thorough description of the process, please see *Chicago region employment forecast: 2021 Update*, available on the CMAP Data Hub after plan adoption in October 2022.

Benchmarking historical employment

Unlike a census of population where every individual is counted as one person, counts of employment are subject to a variety of definitional challenges regarding part-time jobs, self-employed workers, domestic workers, and multiple-job holders. For the purposes of this report, "employment" is primarily based on average annual employment by sector as reported by the Bureau of Labor Statistics' Quarterly Census of Employment and Wages (QCEW), which collects employment and wage data from employers covered under state unemployment insurance programs. To round out employment estimates, two other sources were used: the Railroad Retirement Board to account for railroad workers not counted in QCEW, and the American Community Survey for a count of self-employed workers. Excluded from these estimates are active-duty military, private household workers, and elected officials.

The reallocation of workers counted under Temporary Help Services (colloquially referred to as "temp workers") was informed by the Contingent Worker Supplement to the Census Bureau's Current Population Survey. For land use and travel demand modeling purposes, it's preferable to have temp workers identified by the industries in which they actually work.

Four separate benchmark series of historical (2010–2020) employment by NAICS 2-digit sector were developed to suit different agency purposes:

- Total employment without temp worker reallocation

- Total employment with temp worker reallocation (the series used for reporting in this section)
- Wage and salary employment without temp worker reallocation
- Wage and salary employment with temp worker reallocation (the series used for travel and land use modeling, used in Part 2 of this report)

Forecasting employment

Forecast employment totals for each series are based on May 2021-vintage forecasts produced by Moody's Analytics. The projections that serve as the official regional forecast are Moody's "baseline," where there is an equal probability (50 percent) that the economy will perform better or perform worse over the forecast period. Two alternative scenarios were also produced to illustrate the range of possible outcomes: an "upside" scenario (only a 4 percent probability that the economy would perform better), and a "downside" or slow-growth scenario where the economy underperforms throughout the forecast period. Assumptions behind these scenarios can be found in the appendices in the *Chicago region employment forecast: 2021 update* report.

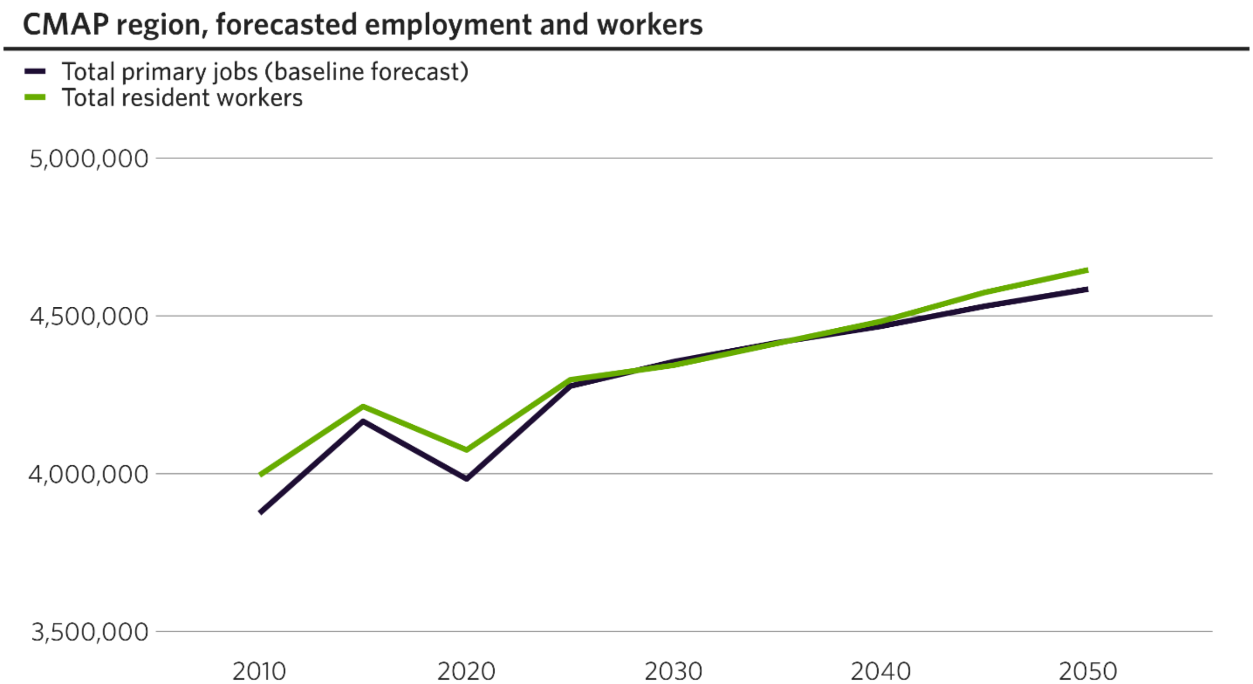
Integration of demographic and economic models

The population and employment models described above operate independent of one another. An additional step is necessary to reconcile labor demand (employment) with labor supply (workers, a subset of the total population). If the rate of employment growth outpaces the number of workers available to fill those positions (through natural increase and baseline net migration), then migration rates are adjusted to address the increased demand for labor. This process primarily affects projections for the working-age population but, as many of these workers are of parenting age, there is a follow-on increase in the youth population as well.

A modified version of Louis Berger's labor-induced migration adjustment model from the ON TO 2050 forecast was used to connect net migration with labor demand. This method retains the assumption that job prospects are a major motivator for people to relocate to northeastern Illinois, a major economic center of the nation. This method also accounts for and excludes additional elements, such as group-quarter populations, non-resident workers, and workers who hold multiple jobs.

For each five-year forecast period, the number of resident workers in the region is estimated by applying a series of modifying factors to the household population: Congressional Budget Office projected labor force participation[5] and unemployment rates,[6] and an out-of-region worker correction factor developed by Berger.[7] Baseline net migration estimated by the demographic model was then modified until the number of estimated resident workers roughly matched the number of primary jobs, within a limited (<3 percent) margin. This estimate of "economic migrants" was allocated across different age groups to retain the known characteristics of net migration in and out of the region (see Migration section above).

Figure 5. Comparison of primary employment and resident workers for the purpose of calculating total net migration



Primary jobs estimated from total employment. Resident workers estimated from projected population, labor force participation rate, and other factors. See the Demographic Forecast Technical Report, ON TO 2050 Plan Update for additional details.

---

[5] Congressional Budget Office, "The 2021 Long-Term Budget Outlook." https://www.cbo.gov/publication/57038

[6] Congressional Budget Office, "An Update to the Budget and Economic Outlook: 2021 to 2031." https://www.cbo.gov/publication/57339

[7] Louis Berger, "Chicago Region Socioeconomic Forecast Final Report" (2016), p. 46. https://datahub.cmap.illinois.gov/dataset/2050-forecast-of-population-households-and-employment

Socioeconomic forecast: Regional results

All results below are for the aggregate, seven-county CMAP region. For sub-regional results please refer to Part 2: Local forecast update. Microsoft Excel versions of all tables will be available on the CMAP Data Hub after plan adoption in October 2022.

Regional population forecast

Forecasts are reported in ten-year intervals for space considerations. A five-year interval version will be available on the CMAP Data Hub.

Table 1. Projected total population, 2020-2050

Total population	2020 (census)	2030	2040	2050
Total population	8,577,735	9,142,057	9,717,333	10,028,854
Non-Hispanic white	4,159,107	4,454,990	4,568,211	4,548,372
Non-Hispanic Black	1,396,303	1,464,567	1,506,422	1,504,683
Non-Hispanic Asian	663,475	807,399	954,695	1,081,180
Non-Hispanic other*	285,541	162,038	182,286	198,255
Hispanic	2,073,309	2,253,063	2,505,718	2,696,364
Percent of total	**2020 (census)**	**2030**	**2040**	**2050**
Non-Hispanic white	48.5%	44.4%	45.6%	45.4%
Non-Hispanic Black	16.3%	15.1%	15.5%	15.5%
Non-Hispanic Asian	7.7%	8.8%	10.4%	11.1%
Non-Hispanic other*	3.3%	1.7%	1.8%	2.0%
Hispanic	24.2%	24.6%	27.4%	29.5%

** Includes: American Indian or Alaska Native (non-Hispanic), Native Hawaiian or other Pacific Islander (non-Hispanic), some other Race (non-Hispanic), and two or more races (non-Hispanic).*

Table 2. Age distribution, 2020 (estimated) and 2050 (projected)

Age group	2020 (census)*	2020 share of total	2050	2050 share of total
0-4	501,945	5.9%	528,877	5.3%
5-17	1,414,967	16.5%	1,501,052	15.0%
18-24	757,136	8.8%	879,221	8.8%
25-44	2,392,250	27.9%	2,353,373	23.5%
45-64	2,198,060	25.6%	2,471,687	24.6%
65-84	1,148,321	13.4%	1,847,042	18.4%
85 & over	165,056	1.9%	447,602	4.5%
TOTAL	8,577,735		10,028,854	

** Due to delays in the release of 2020 Census results, age totals were estimated from 2019 Population Estimates Program county-level results, with proportions applied to 2020 Census Redistricting county-level population totals, then summed to produce regional totals by age group.*

Table 3. Household and group quarters projections

Household population	2020 (census)	2030	2040	2050
Total population in households	8,447,265	8,984,745	9,537,951	9,829,133
Total households*	3,266,741	3,639,601	3,903,663	4,108,756
Average household size	2.59	2.47	2.44	2.39
Group quarters population	**2020 (census)**	**2030**	**2040**	**2050**
Total	130,470	157,312	179,382	199,721
Non-institutional**	67,305	70,940	75,668	80,779
Institutional**	63,165	86,372	103,714	118,942

** Census PL94-171 table H1, "Occupied Housing Units"*

*** Institutional/non-institutional definitions follow Census Bureau designations*

Regional employment forecast

Table 4. Total employment by NAICS sector, 2019-2050

Group	Sector	NAICS	2019	2030	2040	2050
O	Ag., for., fish. and hunt.	11	12,688	11,593	11,157	10,771
O	Mining	21	1,917	1,791	1,735	1,599
T	Utilities	22	14,435	13,240	11,740	10,091
O	Construction	23	219,568	222,393	227,984	237,666
M	Manufacturing	31-33	388,473	350,163	320,543	297,751
T	Wholesale trade	42	221,531	218,183	212,809	200,089
R	Retail trade	44-45	428,246	445,823	448,192	476,170
T	Transp. and wareh.	48-49	287,796	323,390	331,627	326,427
S	Information	51	86,992	81,829	81,996	82,118
S	Finance and insurance	52	230,491	250,217	265,577	282,804
S	Real estate and rental and leasing	53	89,490	90,941	96,694	103,092
S	Professional, scientific and technical services	54	387,388	435,403	473,360	497,008
S	Mgmt. of companies and enterprises	55	68,871	75,316	81,961	86,171
S	Administrative/waste services	56	254,773	267,916	291,992	307,197
S	Educational services	61	400,515	417,264	423,873	424,423
S	Health care and social assistance	62	580,904	610,721	618,001	616,739
S	Arts, entertainment, and recreation	71	125,634	95,573	103,894	117,667
S	Accommodation and food services	72	388,859	329,128	358,042	405,908
S	Other services (exc. pub. administration)	81	198,059	190,196	190,027	193,531
G	Public administration	92	146,532	137,177	134,452	131,891

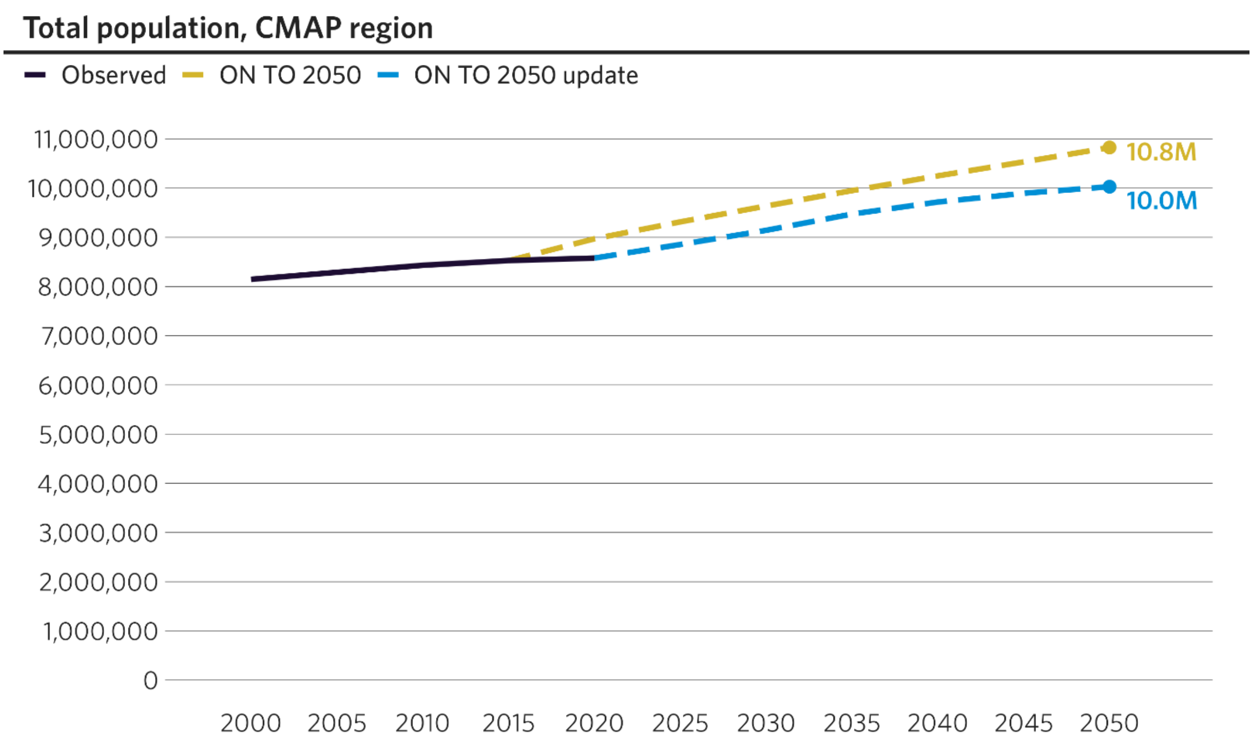
Group	Sector	NAICS	2019	2030	2040	2050
	TOTAL		4,533,162	4,568,258	4,685,656	4,809,114

Comparison to the original ON TO 2050 regional forecast

The new forecast totals for both population and employment are lower than the ON TO 2050 forecast published in 2018. The differences are presented below, followed by an explanation of some of the underlying factors.

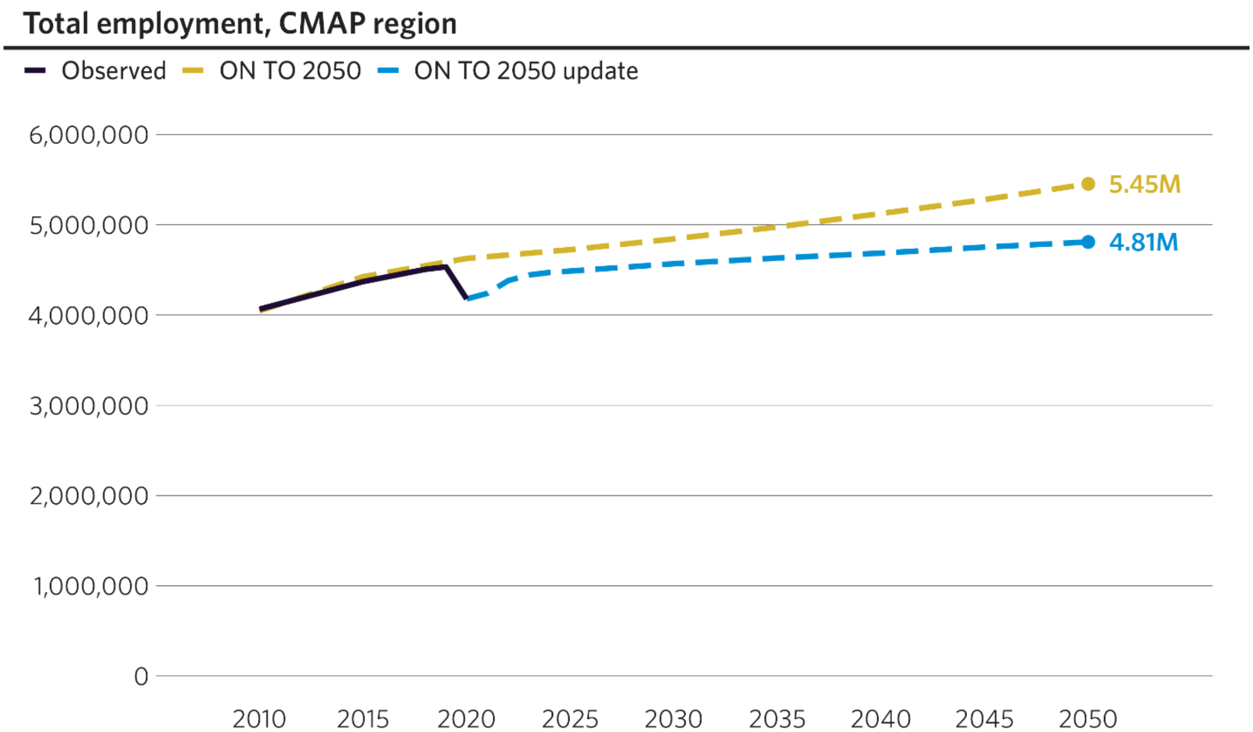
The current population projection of just over ten million persons by 2050 represents a decrease of over 800,000 from the original ON TO 2050 forecast. The graph below shows the region's population growth since 2000 and the divergence of the two forecasts beginning with the year 2015 (the base year for the previous forecast.

Figure 6. ON TO 2050 and plan update population forecasts compared



Similarly, the employment forecast is lower by roughly 600,000 jobs by 2050 from the original ON TO 2050 forecast of over 5.4 million:

Figure 7. ON TO 2050 and plan update employment forecasts compared



Two major contributors to this lowered forecast are lackluster population growth in the previous decade and the impact of the COVID-19 pandemic on regional employment. One additional factor is less obvious: the original ON TO 2050 forecast included a "human capital plus transportation" scenario that we opted not to use in the plan update. That scenario pushed up the previous employment forecast by 175,000 jobs, a number which represents 27 percent of the overall difference between this updated forecast and the original ON TO 2050 forecast; since migration assumptions in the population forecast are based in part on employment projections, there was a follow-on increase in population as well.

Part 2: 2050 Sub-regional forecast update

Introduction

CMAP has invested in the land use microsimulation model UrbanSim to develop local forecasts for the ON TO 2050 update and subsequent plan cycles. Microsimulation models work at a highly disaggregate level (parcels, in the case of UrbanSim), predicting the activities of individual "agents" (households and jobs) over a highly detailed landscape that includes representations of individual buildings, along with known constraints (e.g., zoning) and development events, to simulate land use change in the region over the forecast period. This model has several advantages over the spreadsheet-based Local Area Allocation tool used for ON TO 2050:

- Accounts for local conditions and capacity with parcel-level land use and zoning data
- Creates new residential and non-residential space in a more realistic manner with a developer model
- Allows for more flexible geographic aggregation from a parcel base
- Accommodates complex policy structures for the development of "what if" scenarios

Part 2 will describe the UrbanSim land use model in greater detail, followed by a discussion of model data requirements and the factors employed in the model for the plan update scenario. Sub-regional (county-level plus Chicago) population, households, and employment projections are presented at the end of this section.

The UrbanSim land use model

The UrbanSim model was first developed at the University of California, Berkeley in the late 1990s and has evolved over the years with funding from the National Science Foundation, U.S. Environmental Protection Agency, and the Federal Highway Administration. It is actively used at several metropolitan planning organizations, including the Metropolitan Transportation Commission (California Bay Area), Puget Sound Regional Council (Seattle), the Southeast Michigan Council of Governments (Detroit), and the Metropolitan Council (Minneapolis/St. Paul).

CMAP's implementation of UrbanSim is a cloud-hosted service maintained by UrbanSim, Inc. of Berkeley, California. Model architecture is maintained on GitHub, and model runs are controlled using UrbanCanvas, a browser-based web interface. This document will not attempt to discuss all the workings of UrbanSim but will primarily focus on CMAP data inputs and model enhancements implemented to obtain results for the current forecast cycle. General UrbanSim concepts can be found at https://urbansim.com/urbansim, including documentation of their parcel model.[8]

UrbanSim overview

UrbanSim consists of several sub-models that represent the actions of developers, households, and employers. These include:

- Real estate price model: predicts the per-square-foot rents and prices for each building.
- Real estate developer model: identifies likely locations for new development based on the demand for additional space (forecasted households and employment), allowed uses and densities (zoning), and profitability (prices). Includes a “proforma” model, which evaluates all allowed uses for a site and determines which are likely to be profitable. See Appendix 1 for an overview of the UrbanSim developer model.
- Employment and household transition models: account for new jobs/households (or the loss thereof) in the region, based on regional forecast control totals that determine the number of households by type and employees by industry sector.
- Employment and household relocation models: predict households and employees that may relocate within the region in each model year.
- Household tenure choice model: predicts whether moving households choose to rent or own the housing unit they occupy.
- Employment and household location choice models: predict the location of new (transition models) or relocating (relocation models) households and employment, based on existing available or newly built space.

A graphic representation of these models and their interactions is available in UrbanSim’s online documentation.[9]

Model estimation and calibration

Model estimation is the process of identifying a set of variables that help explain patterns of urban activity, quantifying the relationships of these patterns to cross-sectional variables as known parameters, and determining their relative importance. Calibration involves adjusting these parameters to match longitudinally observed (change over time) data. Estimation and calibration were performed by UrbanSim staff as part of the model development contract.

[8] UrbanSim, Inc. “Parcel-level UrbanCanvas Modeler Documentation.” https://cloud.urbansim.com/docs/parcel-model/modeler-index.html.

[9] UrbanSim, Inc. “UrbanSim Parcel Model.” https://cloud.urbansim.com/docs/general/documentation/urbansim%20parcel%20model.html

Appendix 2 includes entries on the CMAP UrbanSim GitHub Wiki that describe these steps in greater detail.

UrbanSim models can be run using either estimated or calibrated coefficients; early model runs employing calibrated coefficients overemphasized growth in the urban core. This was mitigated by switching to cross-sectionally estimated coefficients in the employment location choice model and assigning a dummy variable to the calibrated household location choice model to de-centralize household growth.

Model data requirements

This section provides a brief overview of the datasets that were collected or created as part of the model development process.

Base-year (2010) datasets

While the model is used to predict growth patterns from the present to the year 2050, the actual modeling period begins at the year 2010 to allow the use of observed (post-2010) trends in model calibration and validation. A detailed description of requirements for these datasets is included in UrbanSim's online documentation.[10]

Table 5. Core base-year (2010) datasets collected for UrbanSim

Requirement	Source	Comment
Parcel geometry	CMAP 2010 Land Use Inventory (LUI) (based on county parcel GIS files)[11]	Many parcels were dissolved into more meaningful "properties" to prevent awkward or unlikely redevelopment of smaller parcels within a larger development.
Parcel attributes	CMAP 2010 LUI, county assessor data	Assessor data were used to obtain land values and tax-exempt status.
Building types	CMAP	List developed to correspond closely with existing LUI land uses (see Appendix 3).
Building footprints	Microsoft (Bing), county GIS files, raster landcover data	Automated process to choose best available source for a given area.
Building attributes	County assessor data, CoStar, commercial websites, LiDAR	Data on building size, value, age, price, and rent. There is no single source for any of these attributes, and many values were estimated or imputed.
Building area per job	U.S. Green Buildings Council, Commercial Buildings Energy Consumption Survey, in-house research	Reported by building type (see Appendix 3) to establish area-per-worker assumptions for forecasted buildings.

[10] UrbanSim, Inc. "Parcel-level UrbanCanvas Modeler Documentation, Data Overview." https://cloud.urbansim.com/docs/general/documentation/urbansim%20parcel%20model%20data.html.

[11] CMAP, "Land Use Inventory for Northeast Illinois, 2010." https://datahub.cmap.illinois.gov/dataset/land-use.

Requirement	Source	Comment
Households/ persons	Census Public Use Microdata Sample (PUMS)	A complete base-year synthetic population was generated using the PopulationSim platform.
Establishments	Dun & Bradstreet (D&B)	Anonymized to sector and job count, geocoded and assigned to buildings based on proximity and known sector/building type relationships.

Regional forecast of households and employment

CMAP's regional forecast (described in Section 1 of this report) is the source of the annual households and employment totals ("controls") throughout the forecast period. As part of the interaction with the agency's four-step travel demand model, household counts need to be broken out by several attributes. Totals generated by the regional demographic model, reported at five-year intervals, were interpolated into annual totals to satisfy UrbanSim requirements.

Table 6. Household control variables

Variable	Description
Number of persons in household	Minimum: 1, Maximum: 7
Age of head of household	Three categories: 16–34, 35–64, 65 and over
Household income	Four categories with breaks at the 25th, 50th, and 75th percentile
Race of head of household	Five categories: Hispanic, Asian non-Hispanic, Black non-Hispanic, white non-Hispanic, other non-Hispanic
16- or 17-year-old in household	Yes/no

Employment control totals relied on the consultant-provided regional employment forecast described in Section 1. Wage and salary, not total employment, is used in UrbanSim for consistency with agency travel model requirements. Five-year totals from the employment forecast were interpolated to provide annual controls required by UrbanSim with one exception: due to the pandemic-related drop in 2020 employment (and rapid near-recovery in subsequent years), employment was interpolated between 2019 and 2025. This was done to limit the number of jobs being removed and then potentially being placed in different locations by UrbanSim in subsequent years. While this provides the model with an inflated number of 2020 jobs, model results for that year are not used by the travel model nor in any reporting of 2020 employment used in this document.

Group quarters (GQ) populations were modeled outside of UrbanSim, using 2020 census PL94-171 block-level counts by the seven major GQ types (institutional: adult correctional facilities, juvenile facilities, skilled nursing facilities, other; non-institutional: college/university student housing, military quarters, other). The increase by GQ type over the forecast period were applied directly at the block level as localized increases in the GQ population. Military quarters population, represented in the region solely by Naval Station Great Lakes in North Chicago, is held constant over the forecast period.

Constraints to development

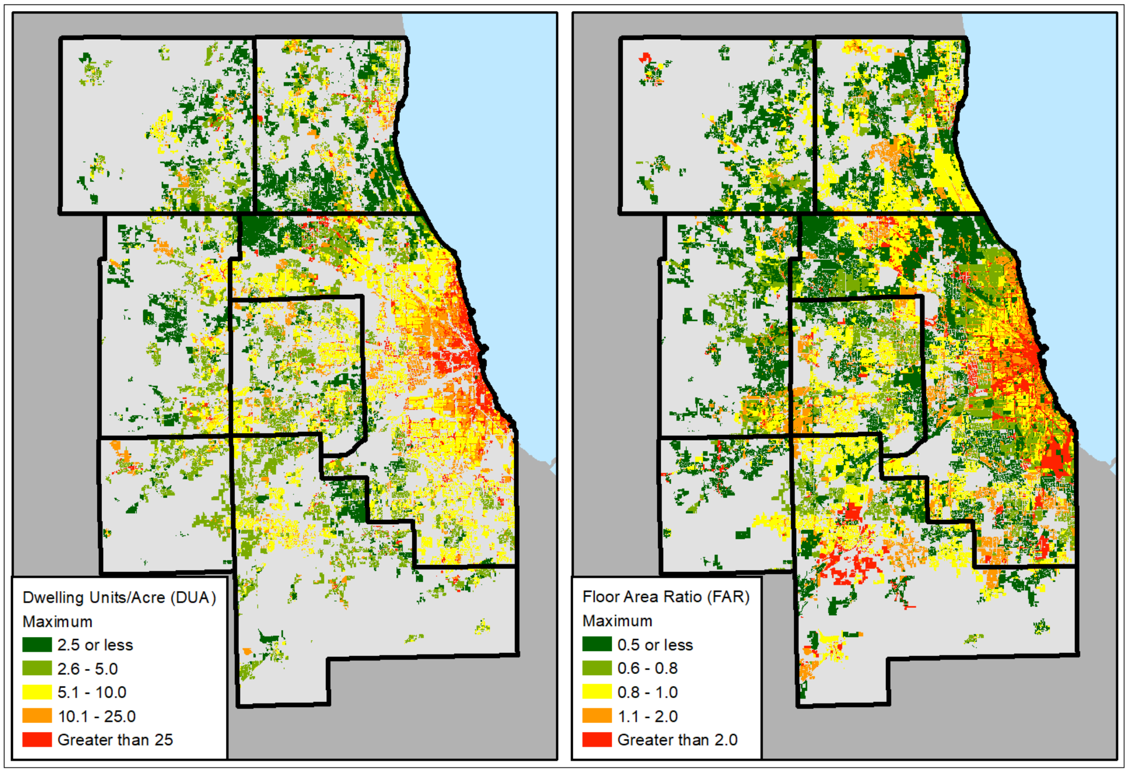
Limits to development are represented in two places in the model. At the parcel level, properties identified as protected open space in the CMAP Land Use Inventory are tagged undevelopable; this was augmented with more recent research to ensure that newly acquired open space was also tagged as undevelopable. Additionally, an undevelopable overlay was created for parcels using FEMA's National Flood Hazard Layer, with a parcel's developable acreage reduced by the percentage of that parcel intersected by that layer.

The second significant constraint on development is local zoning, which limits the type and intensity of development that can occur within each jurisdiction. In the original ON TO 2050 forecast, this concept was articulated through "Urban Classifications," generalized characterizations of development capacity based on existing household and employment density. Working at a more disaggregate level for this cycle allowed us to incorporate zoning, allowing the model to respect community-defined standards for density and use.

CMAP staff undertook a major effort in 2018 and 2019 to collect zoning boundary GIS files (or create them from zoning maps posted online), as well as research local zoning ordinances to identify per-district allowed uses, along with maximum allowable floor area ratio (FAR) and dwelling units per acre (DUA). These data translated directly into the types of buildings that could be allowed on a given parcel, with FAR limiting the size of a non-residential building that could be built on the parcel, and DUA limiting the number of housing units that could be built on the parcel (after subtracting for undevelopable acreage). CMAP was greatly aided in this effort by staff from SB Friedman Development Advisors, who were engaged in a similar effort at the same time. All 284 municipalities in the CMAP region are represented, as well as the unincorporated areas of the seven counties (see **Figure 6**, below).

There are limitations to these data. The collection period was 2018-19, so ordinances updated since that time are not reflected in the model. Special overlay areas, such as historical districts, are not represented. Many ordinances did not explicitly state FAR/DUA values required by the model, so were calculated using other reported restrictions (such as minimum lot size) or inferred from contextual information (similar densities/uses in proximity).

Figure 8. Development constraints: municipal zoning



Known development activity

CMAP's Northeastern Illinois Development Database (NDD) was used to provide UrbanSim with known development activity beyond the 2010 base year, as well as upcoming development projects (the "development pipeline"). NDD records are represented in UrbanSim as individual (or collections of) buildings as an initial step to preempt the developer model from considering or building on those parcels. To conform to model requirements, thousands of NDD records were retrofitted with additional information required for UrbanSim but not collected for NDD. Data used by UrbanSim include status (proposed, committed, under construction, completed), building type, building size (for non-residential buildings), number of residential units, average unit size, and number of affordable units; also, development start year and duration, and flag to indicate whether this is a redevelopment. **Table 7** below lists the post-2020 NDD-derived development assumptions used in UrbanSim.

Table 7. Development pipeline assumptions, 2020 and later, from CMAP's development database (NDD)*

	Residential (units)	Commercial (sq. ft.)	Warehouse/ dist. (sq. ft)	Industrial other (sq. ft)	Other (sq. ft.)
Chicago	28,779	16,185,913	2,069,915	423,454	2,033,234
Suburban Cook	3,212	1,686,414	6,876,592	0	316,064
DuPage	3,420	1,078,833	4,366,554	0	229,124
Kane	1,368	280,725	2,759,146	213,088	3,601
Kendall	228	0	30,742	14,000	0
Lake	393	101,800	2,494,133	0	154,692
McHenry	4,567	28,250	183,000	610,000	358,684
Will	2,849	212,105	8,439,248	906,517	280,421
REGION	**44,816**	**19,574,040**	**27,188,588**	**2,153,059**	**3,375,820**

** Excludes projects that broke ground before 2020*

A separate effort identified larger proposed projects for which we do not have enough information to create an NDD record. This "speculative" project list was researched for the most current available information on proposed types, sizes, and locations of buildings that might be included in each development.

Table 8. Additional development assumptions

Development	Location	Total residential units	Total non-res sq. ft.
One Central	Near South Side (Chicago)	9,240	4,800,000
Lincoln Yards	Lincoln Park/West Town (Chicago)	3,900	1,494,000
Invest SouthWest*	North Lawndale (Chicago)	0	400,000
Bronzeville Lakefront	Douglas (Chicago)	300	595,000
Chicago Bears Move	Arlington Heights	0	4,130,000
AT&T Redevelopment	Hoffman Estates	550	679,683
South Suburban Airport	Peotone	0	850,000

** Represents one Invest SouthWest project whose proposal was not accepted by our deadline. All other Invest SouthWest projects are included in the core development pipeline input.*

All development pipeline data (with attendant unit counts, building size, etc.) are assigned to parcels within the UrbanSim model. After placing these developments (and subtracting new built space from overall new building demand implied by the regional forecast), the remaining building demand is satisfied through the developer model.

Manual adjustments

UrbanSim allows for the manual adjustment of pre-defined areas based on expert opinion and knowledge of the region. This feature is extremely useful in larger institutional settings, which do not conform to behaviors represented in the developer and location choice models. A set of adjustments was identified (Appendix 4) and employment targets created for each area based on existing (2019) employment by industry sector. Employment totals were then forecasted throughout the model period using the per-sector projected increase in regional employment as a guide.

Adjustments for airport and university areas were developed in collaboration with CMAP's travel model team as part of their special generators effort. The proposed South Suburban Airport near Peotone is included in this list, with size and function assumptions based on recent presentations by the Chicago Southland Economic Development Corporation focusing on a cargo airport;[12] the parcel footprint was based on the inaugural configuration in the 2012 Airport Plans Report.[13] The region's two national laboratories (Argonne and Fermilab), as well as Naval Station Great Lakes, were added to this list by the land use team.

Extra-regional model

In addition to the parcel-level microsimulation model developed for the CMAP region, three models representing portions of Illinois, Indiana, and Wisconsin were constructed that correspond with the larger modeling area covered by agency travel models (see **Figure 1** in Part 1). These are aggregate, zone-level models whose base geography is consistent with CMAP travel model subzones.

Data requirements for the zone-based models are less rigorous (e.g., zone-level totals of households and employment, rather than representations of individual households and jobs) and are based on existing development patterns and estimates of additional capacity. This effort is necessary to represent the interaction between the CMAP region and the greater CMAP travel-shed; it is not, however, part of the official CMAP forecast, and results are not presented here. Forecasted household and employment data will be available in the conformity analysis data on the CMAP Data Hub after plan adoption in October 2022. Documentation of UrbanSim's zone-based model can be found on the UrbanSim website.[14]

---

[12] "Column: Amazon growth is driving demand for a third Chicago airport near Peotone, officials say." *Daily Southtown*, 30 April 2021.

[13] AECOM, prepared for the Illinois Department of Transportation. South Suburban Airport Master Plan Draft, 27 September 2012. https://www.southsuburbanairport.com/MasterPlan/reports/ALP/DRAFT_AirportPlansReport-September27-2012.pdf

[14] UrbanSim, Inc. "Zone-level UrbanCanvas Modeler Documentation." https://cloud.urbansim.com/docs/zone-model/modeler-index.html

Policy influence and implementation

This section reviews the market and policy factors used in the original ON TO 2050 local forecast and how those factors were applied in UrbanSim for the forecast update, including two scenario factors unique to UrbanSim's modeling structure that allow for targeted spatial modifications, which align forecast results with ON TO 2050 plan goals.

ON TO 2050 allocation factors and UrbanSim equivalents

As in the original ON TO 2050 forecast, there is a prescriptive element to the updated forecast, meaning it represents one possible outcome if plan recommendations are implemented. The ON TO 2050 forecast had several factors to represent market-based drivers of growth, as well as policy-based "levers" that encourage growth in areas prioritized by the plan, with a result consistent with the goal of encouraging development in infill-supportive areas and reinvesting in disinvested and economically disconnected communities.

The factors used to develop the ON TO 2050 forecast are presented below, with a discussion of how each was addressed in UrbanSim.

Local share of regional households/employment

ON TO 2050 description:

> This factor emphasizes the importance of reinvesting in existing developed areas and incorporating existing densities. Developed areas would be more likely to receive additional residents and employees, and already-dense areas would receive higher amounts (within the prescribed limits of those areas' Urban Classifications).

Base-year population and employment distribution was included as an input to the model estimation process (see Appendix 2).

Local share of overall households/employment over time

ON TO 2050 description:

> This factor builds on the market exhibited by recent growth trends. Prioritizing this factor would emphasize new residents and employment in growing parts of the region.

Observed changes in population and employment between the 2010 base year and latest-available data were used for model calibration.

Property value

ON TO 2050 description:

> This factor serves as an indicator of market potential. Property value depends on many factors, including transportation accessibility, recent development trends,

agglomeration, tax rates, and existing densities. At base, higher property values indicate higher market demand for an area.

Property value, as well as square-foot estimates of prices and rents, are used in UrbanSim developer/proforma models.

Auto/transit accessibility

ON TO 2050 description:

> This factor measures the time required to commute to work from various parts of the region ... Auto and transit accessibility are based on the average generalized cost calculations estimating the average time it takes to travel from one Traffic Analysis Zone (TAZ) to all other TAZs in the region, weighted by population (for the household allocation) and employment (for the employment allocation).

Loading travel model accessibility data (“skims”) directly into UrbanSim replaces this factor, with parcel accessibility inherited from the TAZ that it occupies. Since accessibility evolves over the forecast period due to new or improved transportation facilities and updated population/employment distributions, there is a periodic interaction between UrbanSim and the CMAP travel demand model for the years 2019, 2025, 2030, 2035, 2040, and 2050. Updated household/employment distributions are fed to the travel model, and a new set of skims are uploaded back into the UrbanSim model and used for accessibility estimates in subsequent model years.

Municipal envelope

ON TO 2050 description:

> This factor uses the 2010 municipal boundaries, plus some adjacent area, to allocate growth. GO TO 2040 had a target for 75 percent of new non-residential square footage and 60 percent of new residential units to occur within the 2010 municipal envelope. Prioritizing this factor would emphasize growth in existing incorporated areas.

This concept is manifested in the zoning data which represent development constraints. Unincorporated areas are controlled by county zoning regulations, which generally have lower prescribed densities and more limitations on the types of activities allowed. Incorporated area extent was based on boundaries depicted on each community’s zoning map, or the extent of the shapefile received from the municipality.

Infill supportiveness

ON TO 2050 description:

> CMAP has classified the region into areas with high, moderate, and low potential for infill ... Prioritizing this factor would emphasize reinvestment in existing communities, as well as less-developed areas with municipal plans in place.

The approach to infill for the plan update concentrated on upzoning (increasing the maximum allowed residential density) as the primary tool for encouraging infill development. Increased potential for infill development was directed to outer suburban areas away from sensitive watersheds (as defined by the ON TO 2050 watershed integrity local strategy map[15]), with additional emphasis placed on areas close to Metra rail stations. Final infill lever values are reported in the next section.

Disinvested/economically disconnected areas

ON TO 2050 description:

> Disinvested areas are defined as mature areas that have experienced a combination of population decline, low property values, and high rates of vacancy in residential, commercial, and/or industrial property. Economically disconnected areas (EDAs) contain concentrations of low-income households with either a minority or limited English proficiency population. ON TO 2050 places a priority on renewed public and private investment in these communities. Staff used property value, vacancy, and employment data to identify disinvested areas; assignment of EDAs was based on research in support of the Inclusive Growth strategy paper.[16]

ON TO 2050 disinvested/EDA designations at the census tract level were assigned to all parcels within those tracts.[17] The factor is represented in the UrbanSim proforma (developer) model as an assumption that development costs will be subsidized in these areas, making the proposed development more likely to generate a profit, thus more likely to be built over the forecast period. Final subsidy assumptions are reported in the next section.

Simulation scenario parameters used for the plan update

Unlike the ON TO 2050 local allocation tool, not all the factors discussed above can be articulated as a simple set of values. Exceptions are those factors used as policy levers to encourage development in infill-supportive areas, as well as investing in disinvested and economically disconnected communities. Because of the complex interrelationships among sub-models within UrbanSim, simple one-size-fits-all values for these two levers are not realistic, as they may result in over-building in areas that need little support, and not providing enough support in other areas.

The infill lever used to represent infill supportiveness focused on incorporated areas in outer counties (Kane, Kendall, McHenry, and Will) with existing residential development, away from sensitive watersheds; higher increases were assigned to zoning districts within ½ mile of a

[15] CMAP, "Local Strategy Map: Watershed Integrity." https://www.cmap.illinois.gov/2050/maps/watershed

[16] CMAP, "Inclusive Growth." https://www.cmap.illinois.gov/onto2050/strategy-papers/inclusive-growth

[17] CMAP, "Local Strategy Map: Economically Disconnected and Disinvested Areas." https://www.cmap.illinois.gov/2050/maps/eda

Metra station. Housing unit density (maximum allowed dwelling units per acre, or DUA) was doubled in districts in these areas that allowed single-family detached as the sole residential use. Initial runs incorporating this process resulted in over-building in DuPage and Lake counties; removing the lever from these two counties allowed for more equitable growth among suburban areas.

Table 9. Upzoning levers employed in scenario

Area	Allowed residential types	DUA increase	Start year
Within ½ mile of Metra station	Multiple types	40%	2025
	Single-family detached only	100%	
More than ½ mile from Metra station, 15% of area already developed, not in sensitive watershed	Multiple types	25%	
	Single-family detached only	100%	

The disinvested/EDA subsidy lever was tested and modified (based on interaction with the infill/upzoning lever) to arrive at a subsidy rate that yielded positive growth (in terms of a reasonable increase in jobs and households) in these areas. These subsidies are incorporated in the UrbanSim proforma (developer) model as an assumption that a publicly sponsored subsidy program will reduce development costs in these areas, making the proposed development more likely to be built over the forecast period. The set of census tracts defined as disinvested or economically disconnected in the original ON TO 2050 plan were used for this scenario.

Table 10. Subsidy levers employed in scenario

Area	Subsidy	Start year	End year
Disinvested or disinvested + EDA	3%	2025	2034
EDA alone	2%		
Disinvested or disinvested + EDA	2%	2035	2045
EDA alone	1%		

Socioeconomic forecast: Sub-regional results

Plan updates are intended to incorporate new data and revised assumptions within an existing policy framework, with outcomes that generally reflect those of the original plan. With an updated regional forecast predicting over 800,000 fewer persons and 600,000 fewer jobs by 2050, a corresponding decrease in local forecasts is inevitable. Additionally, the adoption of a land use model for local forecast development means that the distribution of growth won't match the earlier forecast. The spreadsheet tool used in the original ON TO 2050 forecast was indiscriminate in that two places with identical characteristics would be assigned an equal amount of growth; UrbanSim, which creates location-specific developments in response to highly detailed local conditions, will generate different patterns of growth throughout the region.

Local forecast totals

Below are summarized results and maps of the ON TO 2050 Update Local Area Allocation process. Additional data will be made available in October 2022 on the CMAP Data Hub. Census 2020 figures are from the PL94-171 Redistricting file, with the household count based on reported "occupied housing units." 2019 employment estimates are from EBP's benchmark series, with the Chicago share of Cook County's employment based on the city's share of Cook County employment in estimates published by the Illinois Department of Employment Security.[18]

Table 11. Current and projected total population by county and Chicago

	2020 (census)	2030	2040	2050	2020 share	2050 share
Cook	5,275,541	5,565,681	5,860,178	6,016,160	61.5%	59.9%
DuPage	932,877	991,827	1,045,371	1,050,807	10.9%	10.5%
Kane	516,522	566,803	618,878	652,547	6.0%	6.5%
Kendall	131,869	147,715	166,418	192,704	1.5%	1.9%
Lake	714,342	772,156	818,377	832,443	8.3%	8.3%
McHenry	310,229	334,725	374,788	419,425	3.6%	4.2%
Will	696,355	762,379	842,521	887,392	8.1%	8.8%
TOTAL	**8,577,735**	**9,141,286**	**9,726,531**	**10,051,478**		
Chicago	*2,746,388*	*3,138,765*	*3,214,049*	*3,216,869*	*32.0%*	*32.0%*
Suburban Cook	*2,529,153*	*2,426,916*	*2,646,129*	*2,799,291*	*29.5%*	*27.8%*

[18] Illinois Department of Employment Security, "Where Workers Work." https://ides.illinois.gov/resources/labor-market-information/where-workers-work.html

Table 12. Current and projected household population by county and Chicago

	2020 (census)	2030	2040	2050	2020 share	2050 share
Cook	5,195,182	5,472,179	5,752,031	5,894,478	61.5%	59.8%
DuPage	919,059	975,578	1,026,426	1,029,346	10.9%	10.4%
Kane	511,034	560,214	611,118	643,818	6.0%	6.5%
Kendall	131,467	147,221	165,847	192,055	1.6%	1.9%
Lake	694,376	750,501	794,911	807,420	8.2%	8.2%
McHenry	308,386	332,319	371,915	416,139	3.7%	4.2%
Will	687,761	752,607	831,546	875,146	8.1%	8.9%
TOTAL	**8,447,265**	**8,990,619**	**9,553,794**	**9,858,402**		
Chicago	*2,698,875*	*3,086,626*	*3,155,064*	*3,151,094*	*31.9%*	*32.0%*
Suburban Cook	*2,496,307*	*2,385,553*	*2,596,967*	*2,743,384*	*29.6%*	*27.8%*

Table 13. Current and projected total households by county and Chicago

	2020 (census)	2030	2040	2050	2020 share	2050 share
Cook	2,086,940	2,263,483	2,374,380	2,478,534	63.9%	60.3%
DuPage	348,216	392,058	419,758	427,932	10.7%	10.4%
Kane	180,374	213,795	242,297	262,179	5.5%	6.4%
Kendall	43,534	55,599	64,357	76,067	1.3%	1.9%
Lake	253,386	294,469	320,568	332,903	7.8%	8.1%
McHenry	114,282	132,453	153,455	176,411	3.5%	4.3%
Will	240,009	287,728	328,821	354,690	7.3%	8.6%
TOTAL	**3,266,741**	**3,639,585**	**3,903,636**	**4,108,716**		
Chicago	*1,142,725*	*1,202,728*	*1,233,740*	*1,302,281*	*35.0%*	*31.7%*
Suburban Cook	*944,215*	*1,060,755*	*1,140,640*	*1,176,253*	*28.9%*	*28.6%*

Table 14. Current and projected wage and salary employment by county and Chicago

	2019	2030	2040	2050	2019 share	2050 share
Cook	2,616,967	2,650,089	2,723,164	2,803,465	62.8%	63.6%
DuPage	618,700	560,446	567,198	575,172	14.8%	13.1%
Kane	213,345	221,159	224,519	228,566	5.1%	5.2%
Kendall	29,176	35,289	36,173	37,159	0.7%	0.8%
Lake	340,908	340,571	343,329	347,695	8.2%	7.9%
McHenry	97,534	120,002	122,099	123,680	2.3%	2.8%
Will	250,380	265,362	279,463	290,205	6.0%	6.6%
TOTAL	**4,167,010**	**4,192,918**	**4,295,945**	**4,405,942**		
Chicago	*1,382,076*	*1,358,744*	*1,404,343*	*1,452,706*	*33.2%*	*33.0%*

Suburban Cook	*1,234,891*	*1,291,345*	*1,318,821*	*1,350,759*	*29.6%*	*30.7%*

Figure 9. Household density by township and Chicago Community Area, 2020 (census) and 2050 (projected)

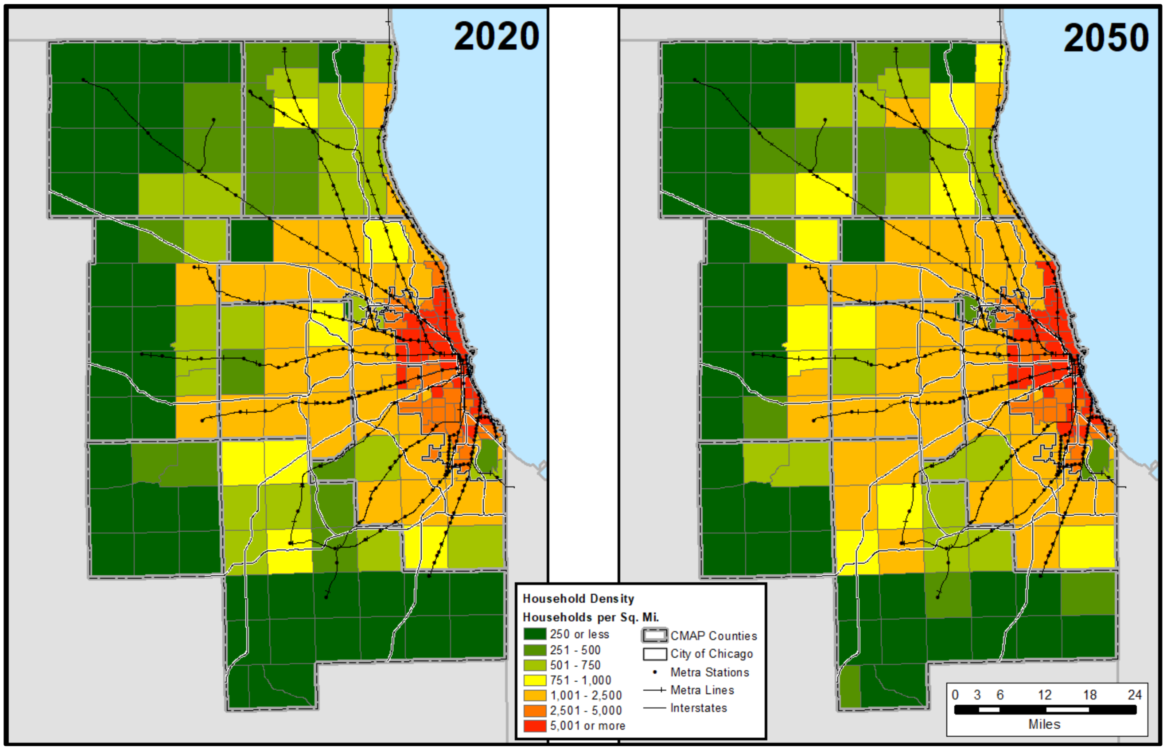
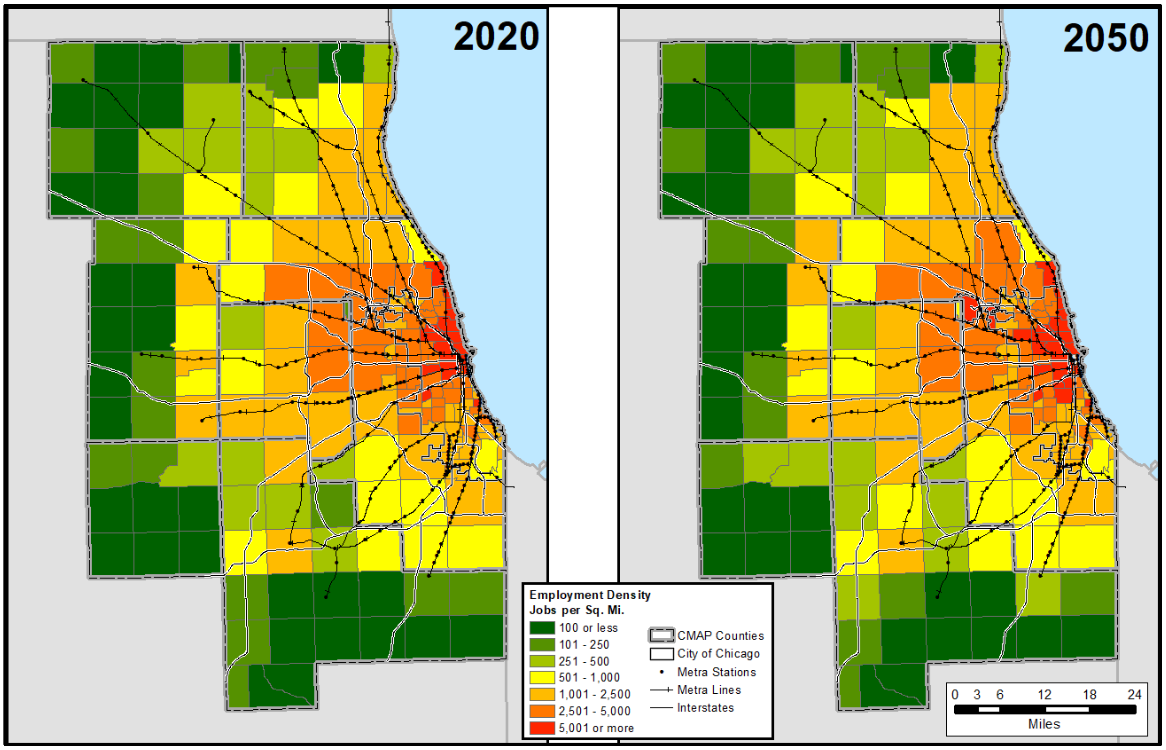


Figure 10. Employment density by township and Chicago Community Area, 2020 (modeled) and 2050 (projected)



Variances with the regional forecast

Some discrepancies will be apparent when comparing results from the local forecast with those from the regional forecast in Part 1. The three contributing factors are discussed below.

Household population: Methodological differences between the regional demographic model (which works at the person level) and UrbanSim (which models entire households) result in discrepancies between the forecasted regional household population in Part 1 and the totals reported here. As both models mature, we will be able to bring these numbers in closer alignment in upcoming forecast cycles.

Group quarters population: 2020 census results were not available in time for inclusion in regional demographic modeling efforts. After the delayed release of the PL94-171 file, the decision was made to incorporate 2020 group quarters counts in the local forecast, with a modification of the scaling process described in Part 1 to accommodate these data.

Employment definitions: Wage and salary employment, described as "workers who receive wages, salaries, commissions, tips, payment in kind, or piece rates [including] employees in both the private and public sectors"[19] is reported in the local forecasts to correspond with

[19] Bureau of Labor Statistics Glossary, "Wage and salary workers." https://www.bls.gov/bls/glossary.htm#W

CMAP travel demand model requirements. Regional forecasted employment in Part 1 reports total employment (wage and salary plus self-employed).

Appendix 1: UrbanSim developer model overview

From the CMAP UrbanSim Wiki on GitHub, by UrbanSim staff. Reprinted with permission.

This page provides a high-level overview of UrbanSim's developer model and its input parameters. The aim is to provide a general description of the process through which the model represents decisions taken by developers in the real estate market and updates the buildings table with extra capacity for every simulation year. Understanding the general logic behind the model, as well as the role of each input parameter, will allow refining the proforma inputs to better represent the context in the CMAP region.

Broadly speaking, the developer model is divided in two steps: feasibility and developer. The feasibility step tests multiple combinations of land use and floor-area-ratio (FAR) for every parcel in the model, returning the most profitable FAR and building configurations for each land use combination in each parcel. This information is then used by the developer step to select the parcels in which new buildings will be built to match existing residential and non-residential demand. A more detailed description of the two steps is given below.

Feasibility step

The feasibility step simulates the typical process that a developer would undergo when deciding what type of development would be most profitable for a given parcel and applies this same logic to all the parcels in the model at a time. The main process can be outlined as follows:

- The proforma is initialized based on the user inputs from the proforma.yaml file, including information about the specific forms that will be tested. Here, each form will represent a combination of land uses that could potentially be built in a parcel (i.e., 80 percent retail, 20 percent residential).
- The sites to analyze and their characteristics are defined based on the parcels table, removing previously pipelined sites.
- For each form (corresponding to a given land use mix):
    - Each potential development site is assigned an acquisition cost that comes from the current yearly rent (either empirical data of rents in the city or forecasts).
    - The model estimates the costs and revenues that would result from building at different alternative densities in the site. (This is done by estimating costs and

revenues that could be obtained from different FARs in each site, with the list of FARs to test being specified by the user inside proforma.yaml).

  - Profit calculations for each potential FAR include the effect of parking requirements, parking costs, building costs at different heights, profit ratio requirements, building efficiency, parcel coverage, and cap rate, among others.
  - Zoning constraints, such as maximum FAR and allowable uses, are taken into account at this point, filtering out those developments that are unfeasible or not allowed. For maximum FAR, the model selects the minimum between the max_far field, and the max FAR that would result from other zoning limits (max heights, max dua, etc).
  - The model generates a feasibility table with the building characteristics that yielded maximum profit for each development site. Building characteristics that make part of the feasibility table include FAR, parking configuration, building sqft, parking ratio, stories, construction time, residential sqft, non-residential sqft, building cost, financing cost, total cost, building revenue, and profit.

The core cost and revenue calculations performed to select the most profitable FAR for each development site for each potential form (land use or land use combination) take place within the Square Foot Proforma API, inside the lookup() function of the feasibility step. The general logic for these calculations is the following:

- Total building area (building bulk) is calculated multiplying FAR by the site area (sqft).
- Building costs are calculated by multiplying built area by cost per sqft for the given building configuration.
- Total construction costs are calculated as the sum of building costs and land costs.
- The loan amount is calculated as total construction costs times loan-to-cost ratio.
- Financing costs are calculated based on the loan amount using the following variables: construction time, drawdown factor, interest rate, loan fees.
- Total development costs are calculated as the sum of construction costs and financing costs.
- To calculate the area (sqft) that will generate rent, common areas and parking are subtracted from the total building area using the parking_sqft_ratio and building_efficiency variables.

- The area that generates rent is multiplied by weighted rent values and divided by the cap rate to calculate the revenue that will be generated by the building.
- Finally, the profit is calculated as the revenue minus total development costs.
- Costs, revenues, and profits are all allowed to be modified by the user through custom callback functions.

One important thing to note is that the feasibility step does all the profit calculations in terms of square footage and has no representation of units (it does not differentiate between rent attained by 1BR, 2BR, or 3BR, and change the results accordingly). Since getting data on unit mixes in the current building stock is extremely difficult, most feasibility computations here happen on a square foot basis, and the developer step handles the translation to units.

Developer step

Having identified the development configuration that would maximize profit for each site-form combination, the main objective of the developer step is to select the sites where buildings will be added on a given simulation year to satisfy demand, and to modify the buildings table to reflect this extra capacity. The main input for the developer step is the feasibility table resulting from the previous step, as well as the demand for residential units and non-residential space on a given simulation year.

For a given simulation year, the developer step can be described as follows:

- The demand for residential units (target_units) is calculated based on the number of forecasted households, the number of existing residential units, and the target vacancy rate. Similarly, the demand for non-residential sqft is calculated based on the number of jobs generated each year, the number of available job spaces, and a target vacancy rate.
- The probability of selecting a given building/development is calculated based on the profit values from the feasibility table. The default function calculates this probability for each site in the feasibility table as the ratio between the profit per unit of area of the site and the sum of profit per unit of area over all feasible sites.
- Using the probability distribution over the potential development sites, the model runs a random function to select specific sites where new developments will be built to meet existing residential demand.

- Both the function to calculate probability based on profit values and the function to select development sites based on the probability distribution can be customized by the user.
- Selected developments are dropped from the feasibility table.
- The buildings table is updated, adding extra capacity in terms of new buildings and new residential units.

Appendix 2: UrbanSim estimation/calibration

From the CMAP UrbanSim Wiki on GitHub, by UrbanSim staff. Reprinted with permission.

Model estimation strategy

This wiki page discusses the approach taken with respect to model estimation. The following UrbanSim models in the CMAP model system required statistical estimation of parameters:

- the household location choice model (multinomial logistic regression)
- the employment location choice model (multinomial logistic regression)
- the hedonic model of real estate prices (ordinary least squares regression)

All estimated coefficients were generated within UrbanSim via Jupyter notebooks. Coefficients are estimated on local CMAP data and not borrowed.

Specification of the location choice models in UrbanSim involves deciding which alternative (i.e., location) characteristics should be considered in the model (i.e., explanatory variables). It also involves determining whether to stratify the estimation by some characteristic of the agents making location choices (i.e., segmentation). Stratification reflects the hypothesis that different groups of agents have different locational preferences. For specifying price models, the modeler decides which observations dataset to use (e.g., buildings), which explanatory variables to use, and how to segment the model into sub-models (e.g., by building type).

Both adding/dropping explanatory variables and changing the model stratification are easy to do in the UrbanSim framework and the notebooks that have been prepared for CMAP. New variables are defined using simple pandas expressions (syntax of the Python pandas library). Each model can be iteratively re-specified and re-estimated quickly during the process of developing a desired model specification. In UrbanSim, the model estimation process is tied closely to simulation. Estimation and simulation both take place within the same code-base and framework. In a properly configured model, simulation can occur right after estimation.

We have variable categories in mind when starting the specification/estimation process (based on hypotheses in the literature), but the specific variables to use depend on local data, review of estimation results (examining coefficient sign, significances, measures-of-fit, and other diagnostics), and an iterative process of trying different specifications.

Variable categories we seek to include in location choice model specifications include real estate characteristics, regional accessibility variables, local accessibility variables, and price. For example, a regional accessibility variable we might try is "employment within 20 minutes auto travel time in the a.m. peak period." This variable would be calculated based on skims from the travel model (stored in the UrbanSim travel_data table). A local accessibility variable we might try is whether there is a school within one mile along the local street network, or retail square footage within a half mile. These kinds of variables would be calculated using the Pandana network accessibility library. In the location choice models, price is a key variable that we try in the specifications. It is hypothesized that, ceteris paribus, households/employment will prefer lower prices (i.e., price will have a negative coefficient), although it is not uncommon in discrete choice models of housing location to find insignificant or even counter-intuitive signs on price variables due to omitted variables that are correlated with price. We also typically include clustering variables. For example, household income interacted with mean income within 400 meters may be tried as an explanatory variable to identify tendencies for income clustering. Similarly, in the employment location choice model, we may try a variable for the number of jobs of the same sector within the zone to capture agglomeration economies.

We start the variable selection process by adding variables to the specification based on behavioral considerations. For example, typical household location choice model explanatory variable categories include:

- price
- residential building characteristics (e.g., year_built)
- neighborhood characteristics
- local and regional accessibility
- interaction variables, such as price interacted with income, or a demographic attribute interacted with a location attribute

Typical employment location choice model explanatory variables include:

- price

- building characteristics (e.g., building type, year_built)
- agglomeration/clustering (e.g., number of jobs within same sector within one mile)
- density (e.g., employment density, population density)
- regional accessibility (skim-based or logsums, e.g. population_within_20_minutes)
- local accessibility (e.g., local street-network based variable)
- composition of households and employment in neighborhood
- If retail-sector, population-seeking variables

Typical real estate price model explanatory variables include:

- distance to local amenities/disamenities
- building characteristics (e.g., year_built)
- regional accessibility (skim-based or logsums, e.g., employment_within_15_minutes)
- neighborhood characteristics (e.g., density, local accessibility, composition)
- Small-area vacancy rates
- Possible (only as needed): geographic dummies for local fixed effects

New variables are defined as python/orca functions in variables.py, and then the variable is added to a model specification using the notebooks, and then the model is estimated and evaluated. We check for fit and significance. If a key behavioral variable (e.g., accessibility) has an intuitive sign but is not significant, we may still retain it for sensitivity reasons.

After trying a set of intuitive behavioral variables, if the model fit is still low, we iteratively try other variables in the specification that have less intuitive interpretations. These less intuitive behavioral variables may be proxying for unobserved factors / unaccounted behaviors, and they help the model to have appropriate spatial associations if behavioral variables alone result in low measures of fit.

For any variable added to a model specification, we consider the resulting metrics:

- Variable significance (t-score)
- Model fit (r2, pseudo-r2)

- Inter-variable correlation matrix to check for multicollinearity (see the plots in the notebook). Correlation coefficients above .6 or so may lead us to reject a variable.
- Variable skew. Excessively skewed variables can result in unreliably estimated parameters. A skew value of greater than 5 or 10 often means we'll try log-transforming the variable to reduce skewnesss.
- Visual assessment of probability plots, or predicted price plots in the case of price models
- If a specification results in a warning being printed about lack of convergence, we make sure to re-run estimation, as the coefficients may not be valid.

Model calibration

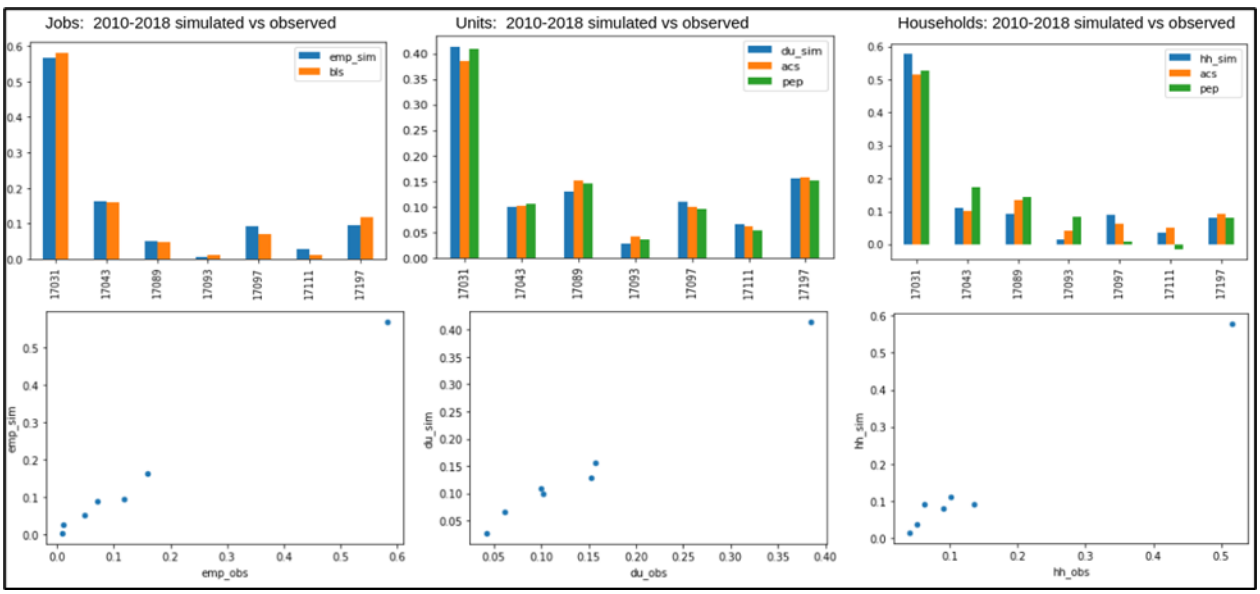
UrbanSim models are estimated using cross-sectional datasets (as represented by the base-year data prepared for the CMAP model); this cross-sectionally-estimated model system is used for forecasting small-area longitudinal patterns of urban growth. “Calibration” in the UrbanSim context means calibrating the temporal dynamics of the simulation to observed longitudinal data. As part of the calibration process, the UrbanSim parameters most likely to have generated observed longitudinal outcomes are inferred. This can help ensure that the model validates better in a forecasting context, a context where the temporal dimension is key.

The approach taken to calibration for the CMAP UrbanSim model is to first frame UrbanSim as a differentiable function, then define an objective function describing the longitudinal accuracy of the model, and then to optimize the objective function by tuning the estimated parameters of the UrbanSim model using automatic differentiation and gradient descent. This can be described as tuning the cross-sectionally-estimated parameters in light of longitudinal data to aid the forecasting accuracy of the model. Note that we are tuning the behavioral parameters of the model, not introducing dummy variables or k-factors to tune; avoiding unnecessary dummies and k-factors is advantageous from the point of view of not dampening the sensitivity of the model system. The model system can then be simulated with either the estimated coefficients or the calibrated coefficients. In the UrbanSim community, simulating based purely on cross-sectionally estimated coefficients is valid, so calibration can be viewed as an optional step. UrbanSim calibration is still an open area of research, and most regions in the past have simulated with cross-sectionally estimated coefficients, but as UrbanSim’s calibration methodologies have been progressing recently, CMAP may choose to take advantage of the calibrate form of their model system.

The goals of model calibration are to move relative spatial variation of simulated growth towards observed patterns (proxy for unobserved costs and variables not accounted for by the models as specified given finite data) and incorporate information from longitudinal data (model estimation is based on cross-sectional data, as mentioned). In both location choice model calibration and proforma calibration, we conduct reverse-mode differentiation (i.e., backpropagation) on the computation graph of the CMAP UrbanSim model to calculate gradients of the scalar-valued loss function (mean-squared error between simulated/observed longitudinal outcomes) with respect to array-valued arguments (the various model input parameters we want to calibrate). We then pass the gradients to an optimizer and do gradient-based optimization to adjust parameter values and minimize the loss.

The calibrated model system was simulated from the 2010 base year to 2018, and then comparisons were made between simulated outcomes and observed outcomes in the 2010-2018 period. This helps to validate the model's performance. The figure below compares simulated with observed outcomes along the employment, residential unit, and household dimensions. The top row contains bar charts of simulated county growth shares compared to observed county growth shares. The bottom row contains scatter plots of the same data. These charts illustrate that the model system's output has a reasonable level of correspondence with observed data on urban growth in the CMAP region.

Figure 11. UrbanSim calibration results



Appendix 3: Building types used in UrbanSim

The table below is a list of all building types represented in UrbanSim, along with square feet-per-job assumptions for all non-residential uses.

Table 15. Building types and utilization assumptions

Building type ID	Building type	Sq. ft. per job	Source
1110	Single-family detached	N/A	
1120	Single-family attached (townhomes / duplexes)	N/A	
1130	Mobile homes	N/A	
1210	Condominium	N/A	
1220	Apartment	N/A	
2110	Mixed use - residential + retail	588	U
2120	Mixed use - residential + office	437	C
2130	Mixed use - residential + services	800	E
3100	Hotel	1,500	E
4000	Storage	8,961	C
4100	Office (Chicago central business district)	300	E
4100	Office	350	E
4210	Grocery store	447	C
4220	Eating and drinking	356	C
4230	Retail - neighborhood	588	S
4240	Retail - strip shopping	758	C
4250	Retail - shopping mall	903	C
4260	Big box retail	826	E
4270	Financial services	305	C
4280	Auto sales / parts / repair	922	C
5100	Light industrial / flex	463	U
5200	Manufacturing	535	U
5300	Warehousing / distribution	1,916	C
6110	Elementary / middle school	1,025	C
6140	High school	1,105	C
6150	College / university	1,003	C
6210	Medical office building	603	C
6230	Hospital	222	C
6300	Parking structures	21,000	E
6400	TCU (transport communication utilities)	2,000	E
6510	Stadium / arena / convention center	1,716	C
6520	Museum	1,884	E
6530	Religious	2,463	C
6540	Other cultural / civic / recreation	1,000	E

Building type ID	Building type	Sq. ft. per job	Source
6610	Administrative / judicial	434	E
6620	Public works / fire / police	256	C
7100	College dormitory	3,127	C
7200	Nursing home	262	C
7300	Military housing	7,818	E
8100	Agriculture	2,500	E
8200	Mining	2,500	E
9000	Other misc.	2,500	E

Sources:

- **C**: U.S. Energy Information Administration, Commercial Buildings Energy Consumption Survey (CBECS), 2012 Public Use Microdata file.[20]
- **E**: CMAP in-house estimate based on CoStar building size/type data with Dun & Bradstreet employment counts.
- **U**: "Building Area per Employee by Business Type," a 2008 document attributed to the U.S. Green Buildings Council but is not available on their website. Originally obtained for CMAP by Louis Berger for the ON TO 2050 forecast.[21]

Note: These sources estimate the number of employees who might occupy a building during peak use. Since we require the total number of employees the building could accommodate over the course of a typical work week, factors were applied to those types which operate beyond a 40-hour week. Examples: retail operations are generally open evenings and weekends; hospitals are continually open (although not necessarily operating at peak). CBECS microdata includes data on operating hours per week, which was used to develop these factors.

[20] U.S. Energy Information Administration, Commercial Buildings Energy Consumption Survey. https://www.eia.gov/consumption/commercial/data/2018/index.php?view=microdata

[21] A copy of this document can be found on the City of Davis (California) website, https://www.cityofdavis.org/home/showpublisheddocument?id=4579

Appendix 4: Adjustment areas

Universities and other institutions with locally controlled employment.

Table 16. Adjustment areas and 2050 employment controls

Adjustment area	2050 employment
O'Hare International Airport	57,377
Midway International Airport	10,636
South Suburban Airport*	334
University of Illinois at Chicago	20,560
Northwestern University	10,886
University of Chicago	23,666
Loyola University	4,684
DePaul University (Lincoln Park and downtown)	6,363
College of DuPage	2,600
Harper College	1,674
College of Lake County	1,363
Moraine Valley Community College	1,223
Naval Station Great Lakes	4,499
Argonne National Laboratory	4,693
Fermilab	2,434
Northern Illinois University (external Illinois model, including surrounding TAZ)	5,061
Gary/Chicago Airport (external Indiana model)	336

** Assumes primary function as air cargo airport*

FY22-0052

433 West Van Buren Street
Suite 450
Chicago, IL 60607

cmap.illinois.gov
312-454-0400

Chicago Metropolitan
Agency for Planning