CITY INTELLIGENCE

GLA Demographic Projections Models

System overview

September 2021

Introduction

Each year, the GLA produces demographic projections to support planning, policy, local service delivery, and more by users across London. To meet the wide range of requirements of its users, the GLA produces dozens of projection variants and scenarios, using alternative input data, assumptions, and methods. To facilitate the design, production, quality assurance, reporting, and dissemination of the many diverse outputs, the GLA uses a specialist suite of demographic models that have been developed over the course of many years.

This document provides an overview of the GLA's demographic projection system, the individual models of which it is currently comprised, how the model is implemented and maintained, and the range of outputs available.

This document does not provide details about specific inputs, assumptions, and configuration choices used in for any particular set of projections. This information can be found in the dedicated documentation accompanying those outputs.

Outline of system

The GLA demographic projection system comprises a collection of individual models intended to be used in combination with one another. The system is designed to provide users with a high degree of flexibility, with a wide range of options available for both the configuration of individual model elements and the interactions between them.

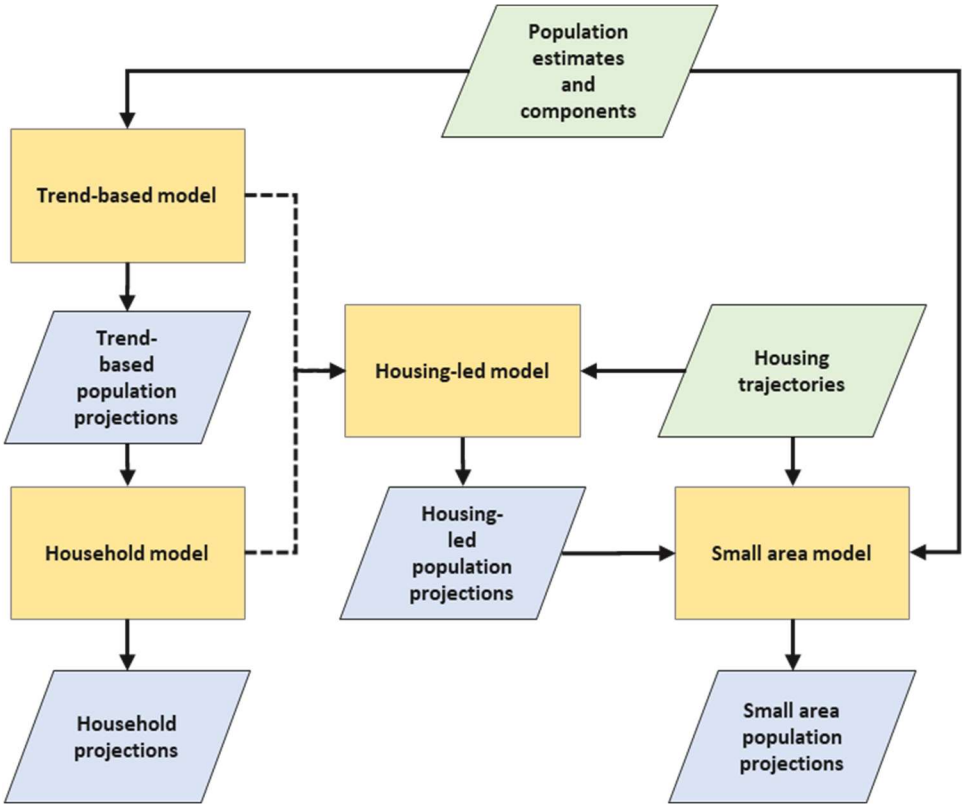
Both the overall system and individual models have been implemented in a manner intended to facilitate their continued development and the inclusion of additional functionality over time.

The current core projection models implemented within the system are:

- *Trend-based population model*
- *Household model*
- *Housing-led population model*
- *Small-area population model*

The conceptual relationship between these four models is shown in the figure below. The models are complimented by a number of ancillary components that provide supporting functionality, such as:

- Preparation of model inputs
- Packaging and dissemination of projection data
- Generation of reporting outputs and visualisations

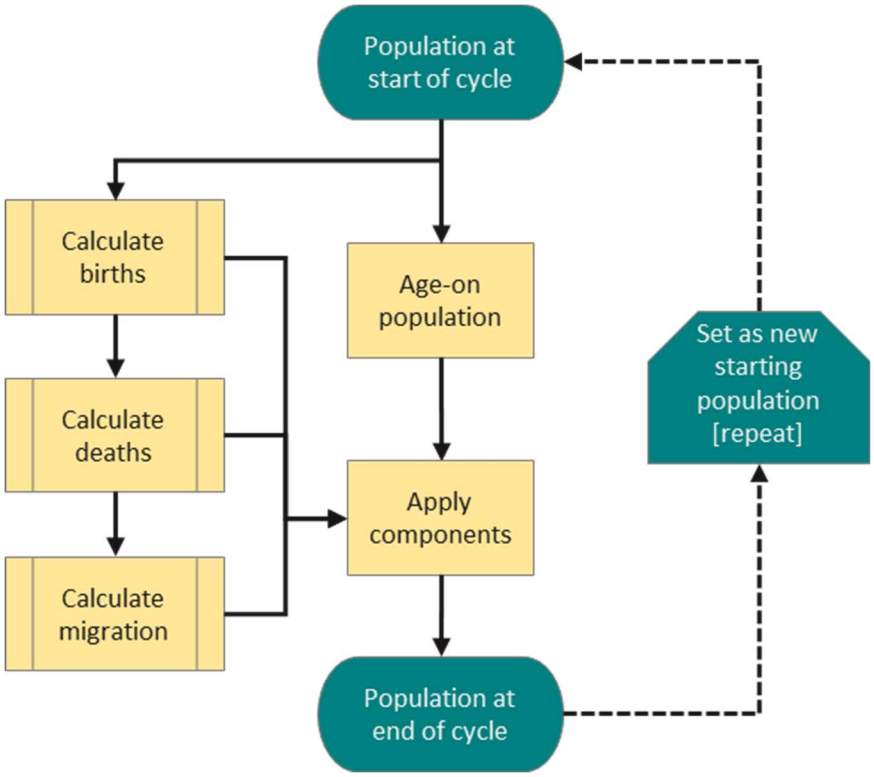


Overview of the core models

The trend-based population model

The GLA's trend-based population model is based on the multi-regional Cohort Component Model (CCM) approach – a standard method employed by statistical agencies around the world to produce national and subnational projections. The cohort component approach is essentially a framework for accounting for population change through the mechanisms of births, deaths, and migration – often referred to as the demographic *components of change*.

CCMs begin with a base population and then roll forward one period at a time (usually, but not always a single year), aging on the population, calculating births, deaths, and migration flows, and then accounting for the impact of these to arrive at the population for the end of the period. In standard implementations of the CCM approach, the calculation of future demographic components is based on a continuation of patterns and trends derived from past estimates. The figure below shows the sequence of calculations.



Such implementations of the cohort component approach don't explicitly account for factors such as housing development or economic growth. However, the past influence of these on population change, particularly through migration, will tend to be reflected in the estimates on which assumed future behaviours are based. Projections based on demographic trends can therefore still reflect implicit assumptions about the future context.

The GLA's trend-based model supports a number of common methods for projecting individual components as well as variations in how these are implemented. Configuration options currently available to the user include:

- Whether to apply "top-down" constraints – forcing the aggregate results for a group of areas to conform with results produced at a higher geographic level.
- The use of rates or fixed flows for projecting international outflows.
- The selection of periods of past data on which to base rates and trends used to project forward.
- Whether to include external assumptions about future changes to fertility or mortality rates.
- Whether to change between alternative sets of migration patterns and behaviours at different points in the projection period.
- Whether to apply user-defined adjustments to the population to account for one-off events.

The default geographic level at which the model operates is Local Authority District (LAD) within England and Wales and national level for Scotland and Northern Ireland. These choices of areas are largely determined by the availability of suitable input data. Local authority districts are the lowest geographical level for which ONS publish annual estimates with detailed components of change for England and Wales; detailed data for areas within Scotland and Northern Ireland are not readily available.

The household model

This model converts a given population into households by removing those living in communal establishments and applying a set of projected household formation rates to the remaining *private household population*.

The household model currently supports two variations on this approach, each based on past official household projections for England. The first was developed to approximate the operation of the model used by Department for Communities and Local Government (DCLG, now MHCLG) for their 2014-based Subnational Household Projections (SNHP)[1]. The second is based upon the model employed by ONS for the 2018-based SNHP[2].

Conceptually, the two models are broadly similar, and the key differences are that:

- The DCLG model used the (now defunct) concept of *head of household*, which defaulted to the eldest adult male in the household, whereas the ONS model uses the current definition of *household representative person* (HRP), who is typically the oldest economically active member of the household.
- They use different periods of past data to project future rates. The DCLG model used census data from 1971 through to 2011, whereas the ONS model was limited to using data from the 2001 and 2011 censuses due to moving to the more modern HRP definition.
- Rates are trended forward for different periods of time. The DCLG model trended forward rates for the duration of the projection period (up to 2039 – the GLA version of this model holds rates constant after this point), while ONS trends rates forward only until 2021 – a design decision necessitated by the short period of past data available on which trends could be based.

The two models tend to give markedly different results for the same input population, with the DCLG model projecting more and smaller households than does the ONS model. This reflects the fact that in many areas, population growth between 2001 and 2011 far outstripped increases in the available dwelling stock, leading to a sharp reduction in household formation among young adults. In the ONS model, this decline in household formation continues at a similar rate for another decade.

While the DCLG model also projects declining household formation among young adults, this is far more gradual due to the trends being based on a much longer period, effectively giving less weight to the change from 2001 to 2011. In addition, the effect of falling household formation among young adults is offset by projected increases in household formation among older adults.

Regardless of which model is used, it is important to note that *all outputs for years after 2011 are projected*, due to the absence of suitable annual household estimates to allow the models to be updated between censuses.

The housing-led population model

The purpose of the housing-led model is to allow projections to be generated that incorporate explicit information and assumptions about past and future housing delivery. This model takes past demographic trends as the starting point for projected future population change, and then modifies these trends based on the estimated capacity of the available housing stock. The number of homes in each year is an exogenous input to the model and can act to constrain or drive additional population growth according to whether its available capacity is judged to be higher or lower than the population projected through a continuation of past demographic trends.

A key challenge in developing models that link population change to housing stock is that the relationship between them is elastic and the result of a complex array of factors. Models that assume too rigid a relationship between population and housing have been shown to perform poorly in London over recent

[1] Detailed information about this model can be found on the government's website here

[2] Detailed information about this model is available from the ONS website here

decades, where population growth has been accommodated through denser occupation of existing dwelling stock as much as through new housing development.

The GLA's current housing-led model is largely the product of pragmatism; its underlying methodology is relatively simple, but it offers the benefits of: modest input data requirements; many years of development and incremental refinement; and a high degree of flexibility in how it is configured.

The housing-led model is an extension of the trend-based population model and it similarly projects forward one year at a time, accounting for population change through the standard mechanisms of births, deaths, and migration. For each projection year:

- An initial trend-based projection of the next year's population is generated.
- The household model is used to estimate the demand for household spaces resulting from this population.
- The demand for space is compared with the estimated capacity of the available housing stock.
- A reconciled 'target' total population is calculated – the value of this target lies partway between that of the initial trend-based projection and that which is consistent with the notional capacity of the housing stock. The position of the target value relative to these two values is determined in the configuration of the model, allowing for simple adjustment of the balance between the influence of past trends and housing capacity.
- Domestic migration flows are adjusted such that the projected population at the end of the year is consistent with the target.

The projected population is therefore influenced by a large number of variables, including:

- Past demographic trends and how the model has been configured to project these forward.
- The absolute and relative level of assumed housing development in each area of the model.
- The setting of the variable that determines the relative influence of past trends and housing capacity on the projected population.
- The choice of household model used to determine the notional demand for housing (the DCLG model results in higher demand for a given population than does the ONS model).
- Whether the model is constrained to match an external projection (in some years, the GLA has opted to constrain the housing-led model for London authorities to a trend-based projection for London as a whole).

While, in principle, the housing-led model can be applied to all areas represented in the trend-based projection model, lack of consistent housing data for areas outside of London has largely limited its use to London local authorities.

The small-area population model

The small-area model is based on a cohort component approach which has been adapted to account for limitations in the migration data available at small geographies. Annual migration flow for England is not currently published below local authority level and so it is not practical to employ a similar multiregional approach to that used in the main trend-based population model.

The model instead is based on a hybrid approach, combining a (simpler) bi-regional cohort component approach with a *housing unit* model. The model uses a combination of dwelling occupation assumptions and origin-destination data from the 2011 census, together with estimated and forecast changes to the available dwelling stock, to generate proxy gross migration flows for each area.

Currently, the reliance on input data from the 2011 census limits the geographies on which the model can operate to 2011 Middle-level Super Output Areas (MSOA) and 2011 Census Merged Wards. While the

model can operate in a standalone manner, it is usually run alongside the main housing-led population model and its results constrained to match at local authority level.

Model outputs

A range of standard outputs are produced as part of each model run, with additional reporting outputs available as configurable options. The range and detail of the available outputs varies between models and is dictated by the specifics of each model's operation. The following sections provide an overview of these outputs.

Trend-based population model

Outputs are produced for each local authority in England and Wales (2021 boundaries). In addition, national-level outputs are produced for Northern Ireland and Scotland. All outputs are available by single-year-of-age (0 to 90+), by sex (female & male), and projection year. Available outputs from the model include:

- Population
- Births
- Births by age of mother
- Age-specific fertility rates
- Deaths
- Age-specific mortality rates
- International In, Out, Net migration
- Domestic In, Out, Net migration
- Population adjustment

Household model

Outputs are produced for all local authorities in England and Wales. No outputs are available for Northern Ireland or Scotland. Outputs are typically only published for London local authorities.

Model outputs are determined by the choice of household model being applied.

DCLG 2014-based household model

- Household population and communal establishment population
    - Sex and age in 18 age bands (5-year bands from 0-4 to 85+)
- Stage 1 household numbers by characteristics of head of household:
    - Sex and age in 18 age bands (5-year bands from 0-4 to 85+)
    - 3 marital status categories: Single, Couple, Previously married
- Stage 2 household numbers by type
    - Age in 9 age bands (10-year bands from 15-25 to 85+)
    - 8 household types:
        - A couple and one or more other adults: No dependent children
        - Households with one dependent child
        - Households with three dependent children
        - Households with two dependent children
        - One family and no others: Couple: No dependent children
        - One person households: Female
        - One person households: Male
        - Other households

ONS 2018-based household model

- Household population and communal establishment population
    - Sex and age in 17 age bands (0-15, 16-19, 5-year bands from 20-24 to 90+).
- Stage 1 households by characteristics of household representative person
    - Sex and age in 17 age bands (0-15, 16-19, 5-year bands from 20-24 to 90+).
- Stage 2 households by type
    - Age in 17 age bands (0-15, 16-19, 5-year bands from 20-24 to 90+).
    - 5 household types:
        - Households with one dependent child
        - Households with three or more dependent children
        - Households with two dependent children
        - One person: Female
        - One person: Male
        - Other households with two or more adults

Housing-led population model

In usual operation, the model is configured to produce outputs for the 33 London local authorities only. Most population outputs and components are available by single-year-of-age (0 to 90+), by sex (female & male), and by projection year. Available outputs from the model include:

- Population
- Private household population (by age band)
- Communal Establishment population (by age band)
- Births
- Deaths
- International In, Out, Net migration
- Domestic In, Out, Net migration*
- Population adjustment
- Number of dwellings
- Number of households
- Average household size

*Note that gross domestic In and Out flows are not available for London as a whole due to the way these components are calculated and applied within the model.

Small-area population model

The small area model requires data inputs from the 2011 census and these inputs dictate the spatial resolution of the model's operation and its outputs. Currently, outputs are available for London census wards (2013 boundaries) and London middle-level super output areas (MSOA). Ward outputs are not published for City of London where ward population sizes are too small to model with any confidence. The ward model outputs for 624 areas. At MSOA-level the outputs are provided for all 985 London MSOAs.

Population outputs are available by single-year-of-age (0 to 90+), by sex (female & male), and projection year. Available outputs from the model include:

- Population
- Births
- Deaths
- Total net migration
- Number of dwellings

Implementation of the system

The GLA's demographic models are modular, and each model combines different methods to construct data inputs and produce projected populations and components of change. Models are implemented in R as a suite of bespoke packages developed and maintained by the Demography team.

The **`popmodules`** package contains functions for running model operations, as well as code for testing and validation, and miscellaneous functions for data wrangling.

The **`trendmodel`**, **`housingledmodel`** and **`smallareamodel`** packages contain functions which leverage **`popmodules`** functions to produce population projections.

Version control and code review are managed through Git and GitHub and all changes to the codebase undergo peer review before being accepted.

Model outputs are produced both in native R format (rds) and machine-readable CSV. Summary Excel outputs are produced using Python scripts called from bespoke R functions. Outputs are disseminated via the London Datastore using the Datastore API[3] and the **`ldndatar`** package developed and maintained by the City Data Team.

[3] https://data.london.gov.uk/guidance/datastore-api/

GLAINTELLIGENCE

For more information please contact: GLA Intelligence

Wil Tonkiss,
Greater London Authority,
City Hall,
The Queen's Walk,
More London,
London SE1 2AA

e-mail: demography@london.gov.uk

Copyright © Greater London Authority, 2020

MAYOR OF LONDON