Population Review

**Volume 58, Number 2, 2019**

**Type: Article, pp. 100-116**

# Comparing Artificial Neural Network and Cohort-Component Models for Population Forecasts

**Authors:** Viktoria Riiman, Amalee Wilson, Reed Milewicz, Peter Pirkelbauer
**Affiliations:** Center for Business and Economic Research, University of Alabama (Riiman); Computer Science Department, Stanford University (Wilson); Center for Computing Research, Sandia National Laboratories (Milewicz); Center for Applied Scientific Computing, Lawrence Livermore National Laboratory (Pirkelbauer)
**Corresponding author/address:** Viktoria Riiman, Center for Business and Economic Research, University of Alabama; email: vriiman@cba.ua.edu

## Abstract

Artificial neural network (ANN) models are rarely used to forecast population in spite of their growing prominence in other fields. We compare the forecasts generated by ANN long short-term memory models (LSTM) with population projections from the traditional cohort-component method (CCM) for counties in Alabama, USA. The evaluation includes projections for all 67 counties, which are diverse in population and socioeconomic characteristics. When comparing projected values with total population counts from the 2010 decennial census, the CCM used by the Center for Business and Economic Research at the University of Alabama in 2001 produced comparable or better results than a basic multi-county ANN LSTM model. Results from ANN models improve when we use single-county models or proxy for a forecaster's experience and personal judgment with potential economic forecasts. The results indicate the significance of forecaster's experience/judgment for CCM and the difficulty, but not impossibility, of substituting these insights with available data.

## Keywords

Population forecast, population projection, artificial neural networks, cohort-component method

**Data availability statement:** The data that support the findings of this study are openly available in OSF at https://osf.io/89WFN/, identifier DOI 10.17605/OSF.IO/89WFN.

**Acknowledgement:** We are grateful for the constructive comments from Ron Prevost. This work was run on computer equipment purchased using NSF grants CNS-0821497 and CNS-1229282.

## 1. Introduction

Artificial neural networks (ANN) are frequently used for forecasting in numerous domains (Crone et al. 2011), such as finance (Niaki and Hoseinzade 2013), biology (Chon et al. 2000) and tourism (Claveria et al. 2015). Few attempts at using them for population estimates and projections, however, have been made thus far (Nordbotten 1996; Tang et al. 2006; Bandyopadhyay and Chattopadhyay 2006; Folorunso et al. 2010). This is despite the fact that the potential for using ANN models in projecting population was noted more than a decade ago (Smith et al. 2001). Prior use of ANN models have utilized feed-forward networks with back propagation (Folorunso et al. 2010; Bandyopadhyay and Chattopadhyay 2006; Nordbotten 1996) or fuzzy networks (Tang et al. 2006). These models were shown to perform better than ratio correlation regression models for projecting population (Tang et al. 2006), as well as for predictions that plug projected fertility, mortality, and migration data into a cohort-component equation (Folorunso et al. 2010). Specifically, Tang et al. (2006) use data on birth, death, and school enrolment to compare their results with the 2000 Census, while Folorunso et al. (2010) use 1990-2060 fertility, mortality, and migration data produced by the National Population Commission to compare their results with target population predictions.

In this paper, we forecast population using a long short-term memory (LSTM) network. LSTM models have become increasingly popular due to their ability to retain memory. We compare projection capabilities from our ANN LSTM models with the population projections developed at the Center for Business and Economic Research (CBER) at the University of Alabama in 2001. CBER researcher, Carolyn Trent, projected population using the cohort-component method (CCM). We assess the accuracy of both methods by comparing them to actual population counts from the 2010 Census or mid-year population estimates by the U.S. Census Bureau.

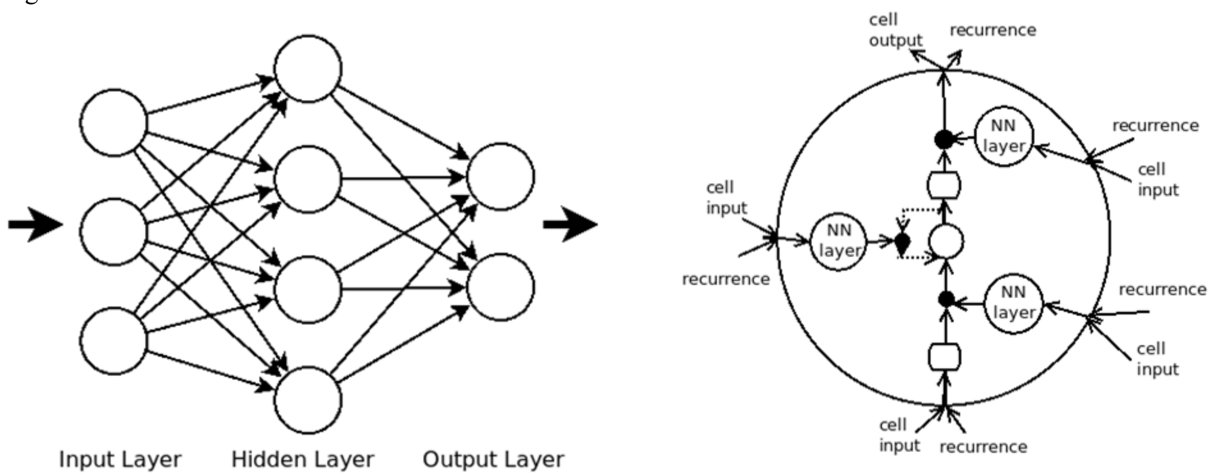
To attain a high-quality cohort-component model, researchers refine the forecasting methods using their experience. ANN models do not have that capability yet, hence we experiment by proxying the lack of cognitive ability with actual economic and demographic data. After experimenting with different types of models and training methods, the results showed that CCM, in general, provided comparable or better results than a basic multi-county ANN LSTM model. Using a single-county ANN LSTM model improved the results overall compared to CCM.

## 2. Artificial neural network model – LSTM

An artificial neural network (ANN) is a machine-learning approach that attempts to simulate cognitive functions. Its characteristics allow for modelling nonlinear relationships between data points. As depicted by Fig. 1a, a simple ANN consists of layers of cells. Each cell receives input from cells of the preceding layer and sends its output to the cells of the following layer. A cell's transfer function computes a weighted combination of its input connection, and fires an output if it exceeds a certain threshold. Before we can utilize an ANN, we need to train it. Under supervised training, input data and expected output data are provided for the model. During the training phase, the ANN adjusts its cells' weights and thresholds to produce the desired output. The most popular technique for that is backpropagation (Folorunso et al. 2010).

Several architectures of neural networks have been proposed (Lipton et al. 2015). The simplest networks are feed-forward networks. Feed-forward networks have been applied to predicting time-series data (Tang and Fishwick 1993; Claveria et al. 2015), including population (Folorunso et al. 2010). A refinement of feed-forward networks, that is, fuzzy networks, has also been applied to population projections (Tang et al. 2006).

Figure 1: Overview of Neural Network




(a) A neural network

(b) A cell in an LSTM network

A drawback of forward networks is the lack of memory, as cells do not have the ability to remember previous outputs. A recurrent network model (RNN) provides the notion of memory. An RNN's cell output of time $t$ is fed back as an input at the time step $t+1$ to the same cell. Thereby, an RNN can retain information for an infinite amount of time. This makes it a powerful model (Siegelmann and Sontag 1995), useful to time-series predictions (Hüsken and Stagge 2003; Claveria et al. 2015).

While RNNs introduce the notion of memory, training a network to retain information for a long time is difficult (Hochreiter 1991; Bengio et al. 1994; Hochreiter et al. 2001). To overcome this challenge, long short-term memory (LSTM) networks have been proposed (Hochreiter and Schmidhuber 1997; Gers et al. 2001). LSTM networks enhance RNN models with the addition of long short-term memory. Thus, the cell can take its history into account. At any given time-step a cell can selectively choose to forget or replace some of the memory. Fig. 1b shows an example of one of the LSTM network cell designs. An LSTM network's input is comprised of the cell's input and the recurrence of its output at the previous time step. The content of LSTM network memory is controlled by internal layers that supervise how much old information is retained and what new information is added. As a result, the output is computed, taking all this information into account.

## 3. Cohort-component method

The cohort-component method is a traditional approach for population projections in demography (Smith et al. 2013; Cheeseman Day 1996; Campbell 1996; Shryock et al. 1973; United Nations 1956; Whelpton

1928). It is also the most popular method among the members of the U.S. Census Bureau Federal-State Co-operative for Population Projections (FSCPP). According to the 2015 FSCPP survey, 75% of FSCPP members use cohort-component method based on historical demographic data (Hunsinger 2015). The next two popular methods were used by 27.5% and 22.5% of respondents, respectively: trend extrapolation of total population data and top-down methods such as constant-share, shift-share, and share of growth.

The Center for Business and Economic Research at the University of Alabama, established in 1930, has a long history of developing population projections using the cohort-component method. The Center researchers use bridged-race population estimates from the Centers for Disease Control and Prevention (CDC), specified in five-year age groups from 0-4 through 80-84 (Center for Business and Economic Research (CBER) 2001). For computational purposes, the 0-4 age group is split into under 1 and 1-4 components, while individuals 85 and over are grouped in a single category. Breaking each age group down by race and sex yields 76 age/race/sex cohorts (using two race groups: white population, as well as black and others population).

The basic equation of the projections is: $P_t = P_{t-1} + B_{t-1,t} - D_{t-1,t} + M_{t-1,t}$
where $P_t$ refers to population at time $t$; $P_{t-1}$ population at time $t-1$; $B_{t-1,t}$ the number of births in the interval from time $t-1$ to time $t$; $D_{t-1,t}$ the deaths in the interval from time $t-1$ to time $t$; and $M_{t-1,t}$ is the net migration in the interval from time $t-1$ to time $t$, which equals in-migration minus out-migration.

The projection process is carried out in five-year increments, run independently for each geography. The calculation of birth, death, and migration components forms the basis for these projections.

For each forecast interval, births are computed as follows: 1) the number of females 15-44 by race in both the current period and the previous period is tallied and these counts are averaged to yield the fertility cohort ($F_{15-44,\,t-1,\,t}$); and 2) expected births by race and sex for a five-year period ($B_{t-1,\,t}$) are calculated as:

$$B_{t-1,t} = 5 * F_{15-44,t-1,t} * FR * SR$$

where *SR* refers to sex ratio calculated from births between the last two decennial censuses and *FR* is general fertility rate, calculated separately for each of the two race groups. Initial fertility rates are calculated as the average of 1989, 1990, and 1991 births by race of the mother divided by the fertility cohort of corresponding race or the number of women aged 15-44 in 1990. Fertility rates for both race groups were held constant throughout the projection period and not converged, which is consistent with the U.S. middle series population projections at the time. Births less deaths before age one become the 0-4 age group for the next projection period.

State and county-specific annual death rates for each race, sex and age group are computed by averaging reported deaths and dividing them by the base population. For all except the 0-4 age group, these one-year death rates ($DR_1$) are converted to five-year rates ($DR_5$) using the equation:

$$DR_5 = 1 - (1 - DR_1)^5.$$

Within the youngest age group, the number of infant deaths is computed as the projected number of births times the calculated probability of death before age one. The death rate for 1-4 year olds is converted to a four-year rate and applied to the remaining population in this group. Death rates are modified for each five-year projection period using the five-year average rate of change from the Social Security Actuarial Life Tables. The largest reductions in mortality over the projection period are seen in the youngest age groups. Mortality is also reduced for individuals 75 and over across the projection period. For cohorts with an average of fewer than five deaths per year, the state death rate for the corresponding cohort is used instead. Age groups with fewer than five deaths will have death rates with large standard errors that will normally include zero within a 95% confidence interval, which implies a true death rate of zero and is viewed skeptically.

Data on migration for the total population of each geography are estimated based on the historical experience, where births (or deaths) from 1990 to 1995 were totaled as:

$$0.75*B_{1990} + B_{1991} + B_{1992} + B_{1993} + B_{1994} + 0.25*B_{1995}$$

And births (or deaths) from 1995 to 2000 were totaled as:

$$.75*B_{1995} + B_{1996} + B_{1997} + B_{1998} + B_{1999} + 0.25*B_{2000}.$$

To calculate migration by age, race, and sex, a survived population was projected for each group in 2000, using the 1990 base population, computed birth rates, and death rates. The difference between this projected population and the actual population of the group in 2000 was then taken to be the residual net migration. Ten-year migration rates were obtained by dividing the residual net migration by the average of the 1990 and 2000 population in each group. These rates were then divided by two to yield five-year migration rates.

For the state and most counties, rates of out-migration were decreased by 10% every five years across the projection period, while rates of in-migration were decreased by 5%. Dampening the rates of in- and out-migration moderates the trends seen between 1990 and 2000. For a few counties experiencing rapid growth, rates of in-migration were further reduced to effectively hold the number of migrants constant. Also, in several counties where population losses in a number of age groups have been large, out-migration rates were reduced by 15% instead of 10%.

Additionally, multiple counties are adjusted for institutional effects, such as colleges, universities, military installations, prisons and nursing homes. Generally, assumption is made that populations in these institutions would not change in size or age distribution throughout the projected period.

Overall, discussions of forecasting, being an art as well as a science, are common among forecasters. Projecting population is noted to be an art that is influenced by scientific techniques and that personal opinions, judgments, experience and outlook are used throughout the process (Guimarães 2014; Daponte et al. 1997). Thus, CBER researcher's experience and personal judgment in making assumptions for birth, death, and migration components are important for the accuracy of projections. Moreover, researcher's opinions are also used when verifying that projected total population and age/race/sex distributions for counties make sense given available information about planned economic developments, potential formation of new school districts, expected changes to prison populations, possible army personnel movements, and all other useful local knowledge.

## 4. Empirical evaluation

*Data description*

For ANN LSTM models, we use data available for all 67 Alabama counties: mid-year intercensal population estimates from 1969, developed by the U.S. Census Bureau. We also use decennial census data by county for each census year between 1910 and 2010, available from the National Historical Geographic Information System. The former is the first year when the data for all 67 Alabama counties are available.

For selected models, we also use births and deaths data from the Center for Health Statistics at the Alabama Department of Public Health. Additionally, economic data are used from the Bureau of Economic Analysis such as proprietors' employment, wage and salary employment, real per capita income, and real average earnings per job. Births, deaths, and economic data as mid-year population are available from 1969. In addition, we use dummies for economic development from 2016 Alabama Workforce Development Councils that divide the state territory into 10 geographically compact regions.

Alabama counties offer diversity in terms of total population, population dynamics, and socioeconomic characteristics. Table 1 provides an overview of data range for:

- total population in 1910 and 2010;
- population growth rates between the decennial censuses;
- births, deaths, and net migration (that is estimated by subtracting births and deaths impacts from change in total population) in 2010; and
- real per capita income and average earnings in 2010.

Since CBER projections using cohort-component method for 2010 were based on data up to and including 2000, in order to have equivalent projections for comparison, we use only the data up to and including 2000 for most ANN LSTM models. In order to try to substitute for forecaster's knowledge on upcoming events affecting population, such as economic and housing developments, in some models we add economic and/or births and deaths data for 2001-2010. Though it is not feasible to have such perfect projections, we added them to check the significance of having these data for forecasters.

*ANN LSTM model specification*

We based our model on a reference implementation for time-series prediction (Brownlee 2016) and implemented it in Keras, a Python package that uses the open-source TensorFlow software library (Géron 2017). All developed models consisted of an LSTM layer (2) and a Dense layer from Keras. While the LSTM layer is responsible for most of the work associated with learning the time series prediction, the Dense layer is responsible for producing a single network output from the LSTM layer's output. While the

Table 1: Data Overview

| | State of Alabama | County - min | County - max | County - median |
|---|---|---|---|---|
| Population 1910 | 2,138,093 | 12,855 | 226,476 | 27,155 |
| Population 2010 | 4,779,736 | 9,045 | 658,466 | 34,339 |
| Population growth: | | | | |
| 1910-20 | 9.8% | -20.3% | 36.9% | 7.8% |
| 1920-30 | 12.7% | -21.5% | 39.2% | 4.9% |
| 1930-40 | 7.1% | -5.0% | 30.7% | 5.2% |
| 1940-50 | 8.1% | -23.2% | 62.8% | -4.8% |
| 1950-60 | 6.7% | -21.5% | 61.0% | -7.1% |
| 1960-70 | 5.4% | -21.7% | 70.4% | 0.6% |
| 1970-80 | 13.1% | -10.4% | 74.3% | 10.6% |
| 1980-90 | 3.8% | -15.0% | 49.9% | -0.7% |
| 1990-2000 | 10.1% | -8.5% | 44.2% | 7.6% |
| 2000-10 | 7.5% | -16.1% | 36.1% | 0.9% |
| Births 2010 | 59,979 | 86 | 8,883 | 419 |
| Deaths 2010 | 47,897 | 90 | 6,773 | 427 |
| Net migration 2010-11 | 3,791 | -2,485 | 3,105 | -53 |
| Real per capita income 2010 | $33,510 | $22,656 | $42,164 | $28,775 |
| Real average earnings per job 2010 | $43,472 | $27,159 | $61,343 | $34,241 |

Source: National Historical Geographic Information System, U.S. Census Bureau, Center for Health Statistics, Alabama Department of Public Health, Bureau of Economic Analysis, and Center for Business and Economic Research, The University of Alabama.

Dense layer is a regular, densely connected network of neurons with a linear activation function, the LSTM layer consists of nonlinear functions. Tests were run using more complicated networks, with more and different layers, but it was found that relatively simple models perform well for our analysis.

In order to explore trade-offs of different approaches to project data, we developed two different variations:

- (1) *Model A, multi-county model*. The question that we attempted to answer with this model is whether we can gain better insights into population dynamics by training a single model from data available for all counties. Under this scenario, input was normalized across all counties, which is necessary for maintaining the relative difference between small and large counties. Once the model was trained, on one county data at a time, it was repeatedly used to predict each single county's output for the projected years.
- (2) *Model B, single-county model.* This is a set of models, one per county, each of which was trained separately on that county data only. All trained models share the same specifications. Thus, each model was specialized for projecting a county's mid-year or decennial census population

independently from other counties' data. Any input data and training data was normalized to values between 0 and 1 for a single county.

Both ANN LSTM models offer various parameters that allowed us to experiment with different setups. Since a neural network is repeatedly trained on the same data over *n* epochs, we experimented with different numbers of epochs up to 200 (using batch size of one). Further, we used different window sizes over a number of past years that are fed into the network to project the population at the next point in time. This allows to account for potential autoregressive processes in population data. Another parameter we experimented with was network size, from four to 32 LSTM cells. We also experimented with different loss functions. The mean absolute error function provided overall best results. We relied on the Keras default activation functions and used the Adam optimizer (Kingma and Ba 2014). We also experimented with different sizes of training and validation sets while keeping test sets for decennial census as 2010 and for mid-year population as 2001-2010 data points.

We experimented with two different input data – decennial census and mid-year annual population data – applying one-step-ahead forecasting approach. For the decennial census data, we trained the model for the period between 1910 and 2000, and then we projected 2010. For these data, we started with a time window of size one, essentially basing the projection on the last measured population 10 years earlier. Increasing the window size produced worse results for most counties. The mid-year population was projected over a 10 year period, one year at a time. Thus, the output/forecast at time step $t_1$ was appended to the input data for the next time step $t_2$. Under this scenario, it could be possible that early projection errors amplify over the 10 year period. For the mid-year population model, we found that a prediction window of five gives good results (though for some specific counties the best results varied). Regarding LSTM cells and number of epochs used for training, the best results for Model A with mid-year population had five LSTM cells and 10 epochs, and with decennial census had four LSTM cells and 100 epochs, while for Model B with both types of population data 16 LSTM cells and 100 epochs provided best results.

## 5. Results

Since previous papers comparing ANN with other methods used for population projections showed the results favorable for ANN, we expected a similar outcome when comparing those with projections from utilizing a cohort-component method in Alabama. When running a basic multi-county model, however, (i.e., ANN LSTM model A) we found that the cohort-component method yields more accurate results when mid-year data are used and comparable results when decennial census data are used (see Table 2). Thus, CCM, forecaster's experience and personal judgment are important for the accuracy of results.

When we used a single county model, ANN LSTM Model B, the results improved. Using decennial census data produced better results for all three errors: root mean-squared error (RMSE), mean absolute error (MAE), and mean absolute percent error (MAPE). Using mid-year population data produced smaller MAPE, but still had larger RMSE and MAE.

Table 2: Comparison of Estimate Errors: Cohort-Component vs ANN LSTM, 2010 Data

| Method | RMSE | MAE | MAPE |
|---|---|---|---|
| ***Cohort-Component Method*** | 5,251 | 3,216 | 6.5% |
| ***ANN-LSTM Model A*** | | | |
| Mid-year population | 17,523 | 11,004 | 16.7% |
| Decennial census population | 5,259 | 3,244 | 6.1% |
| ***ANN-LSTM Model B*** | | | |
| Mid-year population | 7,160 | 3,663 | 6.3% |
| Decennial census population | 4,529 | 2,742 | 5.0% |

Model A: Model is trained on data from all counties, with training process done on one county data at a time.
Model B: Each county has a separate model trained on data from that county only. All 67 counties have the same model specification.
RMSE: root mean-squared error, MAE: mean absolute error, MAPE: mean absolute percent error.

We explored substituting forecaster's experience and personal judgment with true data for births, deaths, and economic data during 2001-2010 period (Table 3). Although it is not feasible to have such accurate data, the experiments offer insight into the importance of these data for projections. Because the data were available since 1969, we used it for projecting mid-year population. The results showed that true economic data improved the results in MAPE from 6.3% to 5.0%, but true births and deaths data increased it to 8.8%, though RMSE decreased. Adding births and deaths to economic data improved RMSE and MAE, but slightly increased MAPE.

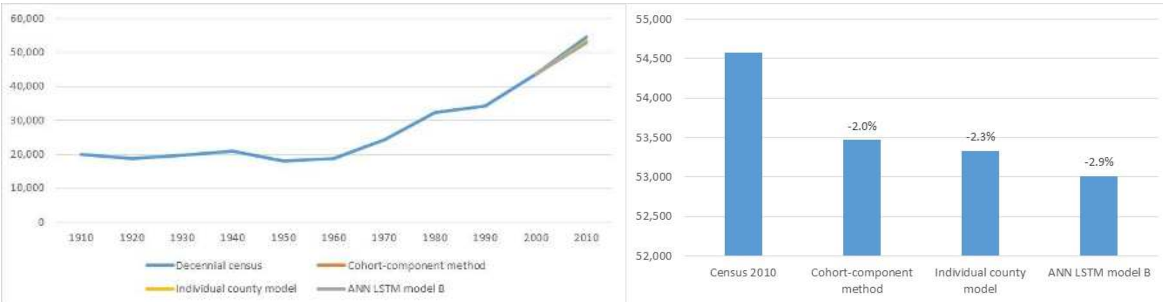
Table 3: Estimate Errors Using True Data in ANN LSTM Model B, 2010 Comparison

| Model | RMSE | MAE | MAPE |
|---|---|---|---|
| True births and deaths | 6,273 | 4,215 | 8.8% |
| True economic data | 5,984 | 3,196 | 5.0% |
| True births, deaths, and economic data | 4,844 | 2,988 | 5.1% |

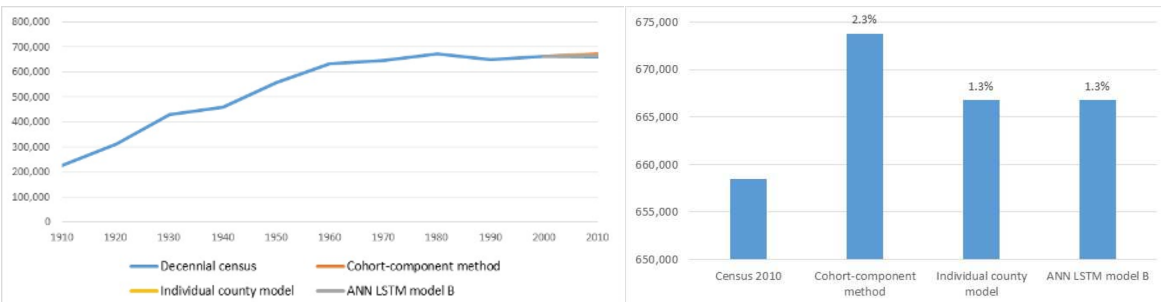
Note: Comparison of projected mid-year 2010 population and 2010 mid-year population estimate from the U.S. Census Bureau.

Since ANN LSTM model B only allows the same specifications for all the counties, we attempted an experiment to see if optimizing a model for an individual county could yield better results. Indeed, choosing specifications to fit a particular county sometimes showed better results than ANN LSTM model B. For Autauga County, for example, a model with individual specification showed smaller errors, with projected population being 2.3% below the true value compared to 2.9% (Fig. 2). The CCM model still produced better results with projections being 2.0% lower than the actual 2010 Census. On the other hand, for Jefferson County, the most populous county in the state, ANN LSTM Model B provided the best results as we did not find any optimization for that county. LSTM model projections were better than the CCM projections for Jefferson County.

Figure 2: Population and 2010 Projections




Autauga County: Population and 2010 Projections




Jefferson County: Population and 2010 Projections

Note: Percentages indicate differences between projections and 2010 Census true values.
Source: National Historical Geographic Information System, U.S. Census Bureau, and Center of Business and Economic Research, The University of Alabama.

We also examined how the models handled the most populous and least populous counties (Fig. A1 and Fig. A2 in Appendix). When looking at the best performing model, ANN LSTM Model B that uses decennial census population data, eight out of 10 most populous counties were among the top 10 counties with largest RMSE and MAE; none were among the top 10 counties with largest MAPE, and two counties were among the bottom 10 counties with smallest MAPE (Table A1 in Appendix). In comparison, when ANN LSTM Model A with intercensal mid-year population data – the worst performing model – was used, five out of the top 10 most populous counties were among the top 10 counties with largest RMSE and MAE, one county was among the top 10 counties with largest MAPE, and two counties were among the bottom 10 counties with smallest MAPE. CCM results showed more counties in the top 10 most populous counties among the bottom 10 counties with smallest MAPE than either of ANN LSTM models (four counties).

## 6. Discussion and conclusion

A forecaster's experience and personal judgment seem to have a strong impact on the accuracy of population projections since results from the cohort-component method were either comparable or more accurate than from the ANN LSTM model A, which was trained on data from all counties, one county data at a time. Results from ANN LSTM model B, which used single county data when training the model, gave better results than the CCM when decennial census data were used. When mid-year population data were used for model B, it gave better MAPE but worse RMSE and MAE than the CCM. Thus, we may still need to find ways to substitute for forecasters' personal experience, judgment, and information available to them when developing population projections. Using more data, such as economic forecasts, could be one such option.

Training ANN LSTM model only on the data from the county for which projections are later made gave better results, indicating that population development trends in other counties did not affect a specific county's population projection as much as a county's own historical trends and instead may have created additional noise for the ANN LSTM model. The diversity of population dynamics among Alabama counties may have been the reason for this noise. Adding more geospatial data to this model could improve the results, capturing the potential impact of geographic location on population growth. Although our experiments with adding economic development region dummies to model A did not improve the results, further exploration is needed.

Using decennial census data for ANN LSTM models resulted in smaller errors than in models with mid-year population. Thus, having a longer time span as input produced better results. This could be caused by the length of projections that required 10 steps for mid-year population compared with one step for decennial census data due to the nature of one-step-ahead forecasts. Using other forecasting techniques, such as direct prediction (Bontempi 2008), could improve projections of mid-year population.

In this work, we only experimented with a small set of parameters of a reference implementation. It is our goal to explore automated techniques for finding better optimized ANN LSTM models (Goodfellow et al. 2016). With the continuing development of ANN models, we are expecting to receive improved forecasts.

Overall, in the future, it could be worthwhile to explore using ANN models to project only some population components instead of total population. This could make ANN models another alternative in the toolbox of demographers.

## References


Bandyopadhyay, G. and S. Chattopadhyay (2006). "An artificial neural net approach to forecast the population of India." Cornell University Library, Nonlinear Sciences. [https://arxiv.org/abs/nlin/0607058, accessed March 7, 2019]

Bengio, Y., P. Simard, and P. Frasconi (1994). "Learning long-term dependencies with gradient descent is difficult." *Trans. Neur. Netw.* 5(2): 157-166.

Bontempi, G. (2008). "Long term time series prediction with multi-input multi-output local learning." In *2nd European Symposium on Time Series Prediction (TSP), ESTSP'08*: 145-154.

Brownlee, J. (2016). "Time series prediction with LSTM recurrent neural networks in python with keras." [http://machinelearningmastery.com/time-series-prediction-lstm-recurrent-neural-networks-python-keras/, accessed December 19, 2016]

Campbell, P. R. (1996). *Population projections for states by age, sex, race, and Hispanic origin: 1995 to 2025.* Technical report. U.S. Bureau of the Census, Washington, DC.

Center for Business and Economic Research (CBER) (2001). *Alabama Population Projections 2010-2040. Methodology.* The University of Alabama, Culverhouse College of Commerce.

Cheeseman Day, J. (1996). *Population projections of the united states by age, sex, race, and Hispanic origin: 1995 to 2050.* Technical report. U.S. Government Printing Office, Washington, DC.

Chon, T.-S. et al (2000). "Use of an artificial neural network to predict population dynamics of the forest–pest pine needle gall midge (diptera: Cecidomyiida)." *Environmental Entomology* 29(6):1208.

Claveria, O., E. Monte, and S. Torra (2015). "Multiple-input multiple-output vs. single-input single-output neural network forecasting." Working paper 2015/02 1/28. Research Institute of Applied Economics, Regional Quantitative Analysis Research Group.

Crone, S. F., M. Hibon, and K. Nikolopoulos (2011). "Advances in forecasting with neural networks? empirical evidence from the {NN3} competition on time series prediction." *International Journal of Forecasting* 27(3): 635 - 660.

Daponte, B. O., J.B. Kadane, and L.J. Wolfson (1997). "Bayesian demography: projecting the Iraqi Kurdish population, 1977-1990." *Journal of the American Statistical Association* 92(440): 1256-1267.

Folorunso, O. et al (2010). "Population prediction using artificial neural network." *African Journal of Mathematics and Computer Science Research* 3(8): 155-162.

Géron, A. (2017). *Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. O'Reilly Media, 1st edition.

Gers, F., D. Eck, and J. Schmidhuber (2001). “Applying lstm to time series predictable through time-window approaches.” In *Proceedings of the International Conference on Artificial Neural Networks*, ICANN '01: 669-676. London, UK. Springer-Verlag.

Goodfellow, I., Y. Bengio, and A. Courville (2016). *Deep Learning*. MIT Press. [http://www.deeplearningbook.org, accessed March 7, 2019]

Guimarães, R. R. d. M. (2014). “Uncertainty in population projections: the state of the art.” *Revista Brasileira de Estudos de Populaçõ* 31: 277-290.

Hochreiter, S. (1991). *Untersuchungen zu dynamischen neuronalen Netzen.* Diploma thesis, Institut für Informatik, Lehrstuhl Prof. Brauer, Technische Universität München.

Hochreiter, S. et al (2001). “Gradient flow in recurrent nets: the difficulty of learning long-term dependencies.” In Kolen, J. and Kremer, S., editors, *Field Guide to Dynamical Recurrent Networks*. IEEE Press.

Hochreiter, S. and J. Schmidhuber (1997). “Long short-term memory.” *Neural Comput.* 9(8): 1735-1780.

Hunsinger, E. (2015). *FSCPP Members 2015 Questionnaire.* Alaska Department of Labor and Workforce Development, State Demographer's Office.

Hüsken, M. and P. Stagge (2003). “Recurrent neural networks for time series classification.” *Neurocomputing* 50: 223-235.

Kingma, D. P. and J. Ba (2014). “Adam: A method for stochastic optimization.” *CoRR*, abs/1412.6980.

Lipton, Z. C., J. Berkowitz, and C. Elkan (2015). “A critical review of recurrent neural networks for sequence learning.” *CoRR*, abs/1506.00019.

Niaki, S. T. A. and S. Hoseinzade (2013). “Forecasting S&P 500 index using artificial neural networks and design of experiments.” *Journal of Industrial Engineering International* 9(1): 1.

Nordbotten, S. (1996). “Neural network imputation applied to the Norwegian 1990 population census data.” *Journal of Official Statistics* 12(4): 385-401.

Shryock, H. S., J.S. Siegel, and associates (1973). *The Methods and Materials of Demography*. A United States Dept. of Commerce publication. U.S. Bureau of the Census, second printing edition. Library of congress number: 76609905.

Siegelmann, H. and E. Sontag (1995). “On the computational power of neural nets.” *Journal of Computer and System Sciences* 50(1): 132-150.

Smith, S. K., J. Tayman, and D. Swanson (2001). *State and Local Population Projections. Methodology and Analysis*. Kluwer Academic/Plenum Publishers, NY.

Smith, S. K., J. Tayman, and D. Swanson (2013). *A Practitioner's Guide to State and Local Population Projections*. Springer Media Dordrecht.

Tang, Z. and P.A. Fishwick (1993). “Feed-forward neural nets as models for time series forecasting.” *ORSA Journal of Computing* 5(4): 374-386.

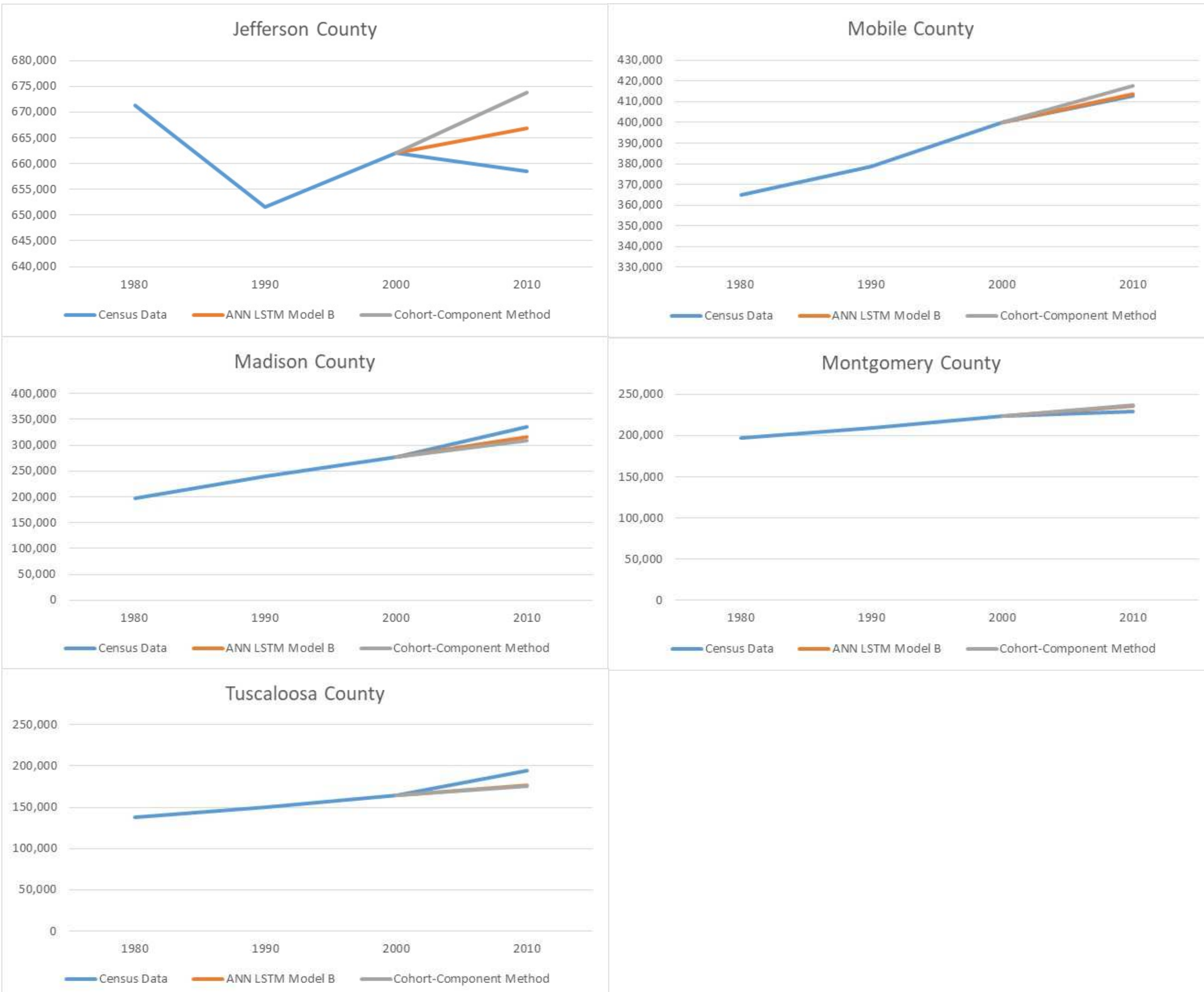
Tang, Z., C.W. Leung, and K. Bagchi (2006). “Improving population estimation with neural network models.” In *Proceedings of the Third International Conference on Advances in Neural Networks - Volume Part III*, ISNN'06: 1181-1186. Berlin, Heidelberg. Springer-Verlag.

United Nations (1956). *Methods for population projections by sex and age*. United Nations, N.Y. Sales no. 56.XIII.3.

Whelpton, P. K. (1928). “Population of the united states, 1925 to 1975.” *American Journal of Sociology* 34(2): 253-270.

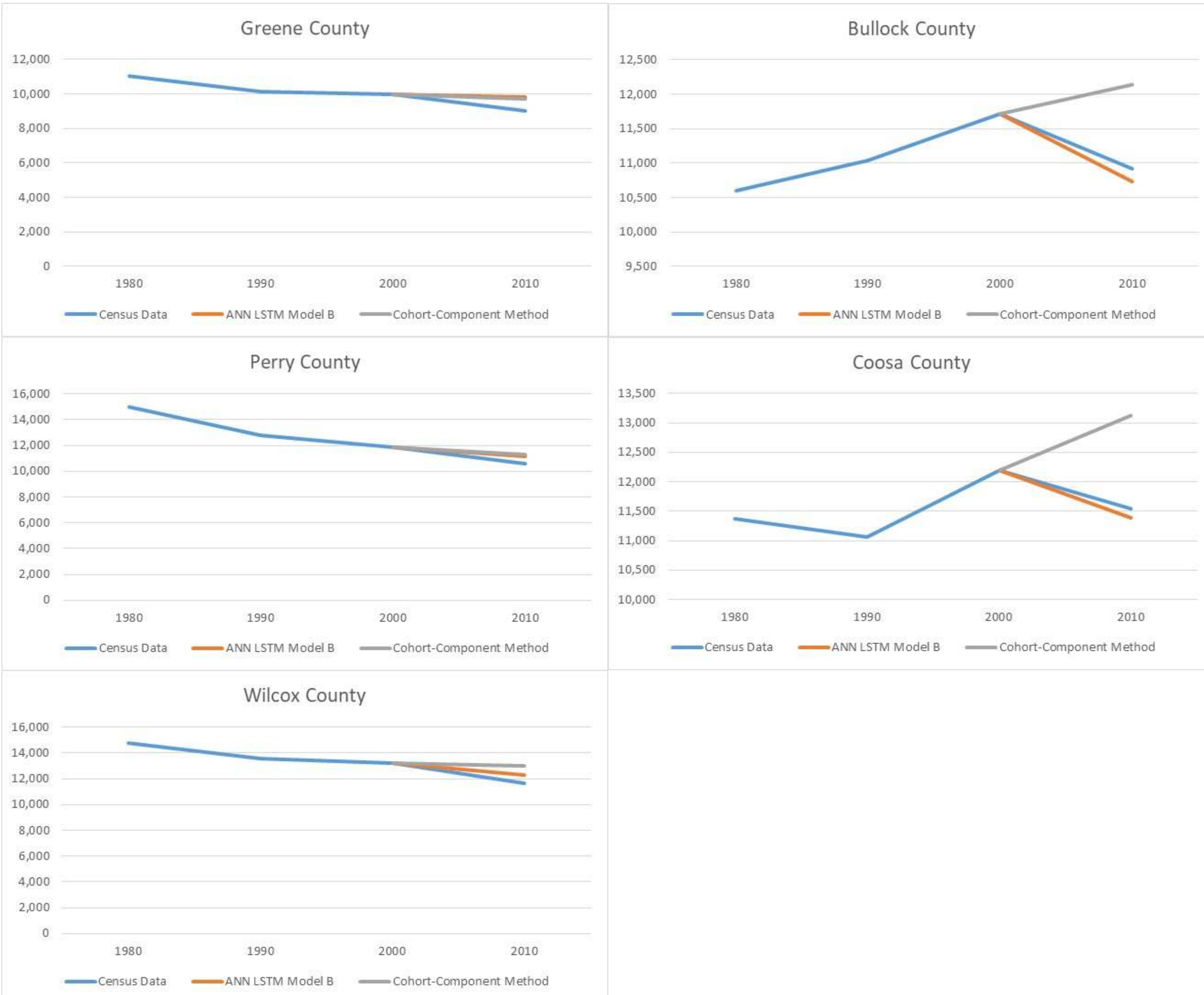
## Appendix

Figure A1: Overview of the Five Most Populous Counties in Alabama




Note: Counties ranked by their 2000 Census population.
Source: U.S. Census Bureau and Center of Business and Economic Research, The University of Alabama.

Figure A2: Overview of the Five Least Populous Counties in Alabama




Note: Counties ranked by their 2000 Census population.
Source: U.S. Census Bureau and Center of Business and Economic Research, The University of Alabama.

Table A1: Error Rankings by County from the Best and Worst Performing ANN LSTM Models

| County | 2010 Census population, rank | Cohort-Component Model | | ANN LSTM Model B, decennial census population | | ANN LSTM Model A, mid-year population | |
|---|---|---|---|---|---|---|---|
| | | RMSE & MAE, rank | MAPE, rank | RMSE & MAE, rank | MAPE, rank | RMSE & MAE, rank | MAPE, rank |
| Jefferson | 1 | 3 | 51 | 5 | 59 | 11 | 63 |
| Mobile | 2 | 12 | 61 | 41 | 67 | 2 | 36 |
| Madison | 3 | 1 | 27 | 1 | 27 | 16 | 59 |
| Montgomery | 4 | 4 | 45 | 9 | 44 | 17 | 52 |
| Shelby | 5 | 16 | 55 | 7 | 36 | 1 | 8 |
| Tuscaloosa | 6 | 2 | 13 | 2 | 11 | 5 | 31 |
| Baldwin | 7 | 32 | 60 | 4 | 30 | 3 | 11 |
| Lee | 8 | 52 | 63 | 3 | 12 | 4 | 12 |
| Morgan | 9 | 56 | 64 | 17 | 46 | 12 | 29 |
| Calhoun | 10 | 8 | 38 | 8 | 24 | 18 | 41 |
| Crenshaw | 58 | 65 | 58 | 56 | 41 | 54 | 43 |
| Choctaw | 59 | 33 | 4 | 29 | 4 | 43 | 18 |
| Sumter | 60 | 67 | 66 | 51 | 32 | 47 | 23 |
| Conecuh | 61 | 54 | 32 | 43 | 21 | 48 | 21 |
| Wilcox | 62 | 47 | 8 | 52 | 28 | 40 | 7 |
| Coosa | 63 | 38 | 5 | 67 | 58 | 49 | 19 |
| Lowndes | 64 | 25 | 1 | 32 | 2 | 35 | 3 |
| Bullock | 65 | 50 | 7 | 66 | 52 | 42 | 10 |
| Perry | 66 | 58 | 33 | 54 | 29 | 37 | 2 |
| Greene | 67 | 59 | 30 | 46 | 13 | 34 | 1 |

Note: The results are from the best performing (Model B, decennial census data) and worst performing model (Model A, mid-year intercensal data) for ten most populous and ten least populous counties. Errors are ranked from the largest (rank #1) to the smallest error (rank #67).

Source: U.S. Census Bureau and Center of Business and Economic Research, The University of Alabama.