LBNL-57955

ERNEST ORLANDO LAWRENCE BERKELEY NATIONAL LABORATORY

U.S. Regional Energy Demand Forecasts Using NEMS and GIS

Jesse A. Cohen, Jennifer L. Edwards, and Chris Marnay

Environmental Energy
Technologies Division

July 2005

http://eetd.lbl.gov/ea/EMS/EMS_pubs.html

This work described in this report was funded by the Office of the Assistant Secretary of Energy for Energy Efficiency and Renewable Energy, Office of Planning, Budget, and Analysis of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

Disclaimer

LBNL-57955

U.S. Regional Energy Demand Forecasts Using NEMS and GIS

Prepared for the
Office of Planning, Budget, and Analysis
Assistant Secretary for Energy Efficiency and Renewable Energy
U.S. Department of Energy

Principal Authors

Jesse A. Cohen, Jennifer L. Edwards, and Chris Marnay

Ernest Orlando Lawrence Berkeley National Laboratory
1 Cyclotron Road, MS 90R4000
Berkeley CA 94720-8136

July 2005

The work described in this paper was funded by the Assistant Secretary of Energy for Energy Efficiency and Renewable Energy, Office of Planning, Budget, and Analysis of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

Abstract

The National Energy Modeling System (NEMS) is a multi-sector, integrated model of the U.S. energy system put out by the Department of Energy's Energy Information Administration. NEMS is used to produce the annual 20-year forecast of U.S. energy use aggregated to the nine-region census division level. The research objective was to disaggregate this regional energy forecast to the county level for select forecast years, for use in a more detailed and accurate regional analysis of energy usage across the U.S.

The process of disaggregation using a geographic information system (GIS) was researched and a model was created utilizing available population forecasts and climate zone data. The model's primary purpose was to generate an energy demand forecast with greater spatial resolution than what is currently produced by NEMS, and to produce a flexible model that can be used repeatedly as an add-on to NEMS in which detailed analysis can be executed exogenously with results fed back into the NEMS data flow. The methods developed were then applied to the study data to obtain residential and commercial electricity demand forecasts. The model was subjected to comparative and statistical testing to assess predictive accuracy. Forecasts using this model were robust and accurate in slow-growing, temperate regions such as the Midwest and Mountain regions. Interestingly, however, the model performed with less accuracy in the Pacific and Northwest regions of the country where population growth was more active. In the future more refined methods will be necessary to improve the accuracy of these forecasts.

The disaggregation method was written into a flexible tool within the ArcGIS environment which enables the user to output the results in five year intervals over the period 2000-2025. In addition, the outputs of this tool were used to develop a time-series simulation showing the temporal changes in electricity forecasts in terms of absolute, per capita, and density of demand.

Table of Contents

List of Figures

List of Tables

Abbreviations and Definitions

AEO – Annual Energy Outlook
ArcGIS - ESRI Geographic Information System software
BTU – British Thermal Unit
CBECS – Commercial Building Energy Consumption Survey
DER – distributed energy resources
DOE – Department of Energy
EIA – Energy Information Administration
EPA – Environmental Protection Agency
GUI – Graphic User Interface
LBNL – Earnest Orlando Lawrence Berkeley National Laboratory
NEMS – National Energy Modeling System
NOAA – National Oceanographic and Atmospheric Administration
Quad – Quadrillion megawatts
RECS – Residential Energy Consumption Survey
VBA – Visual Basic for Applications

Acknowledgments

The work described in this report was funded by the Assistant Secretary of Energy Efficiency and Renewable Energy, Office of Planning, Budget, and Analysis of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. The authors would like to thank our Program Manager, Scott Hassell. Also, a special thank you to Liz Brady Sabeff and Ray George of the National Renewable Energy Laboratory for providing a thorough and insightful review of this report. The authors would also like to thank Judy Lai for her help editing this work.

1. Introduction

The United States (U.S.) Energy Information Administration's (EIA) National Energy Modeling System (NEMS) is a multi-sector, integrated model of the U.S. energy system. NEMS is used by EIA to produce the Annual Energy Outlook (AEO), their official annual 20-year forecast of U.S. energy use. Lawrence Berkeley National Laboratory's (LBNL) ongoing research on distributed energy resources (DER) relies heavily on these forecasts to perform various types of energy analyses to predict the spread of alternative energy sources through distributed generation (local generation sources sited at the point of end use), to estimate the effects of proposed energy efficiency standards, to predict changes in interregional electric grid congestion, and in other analyses. The AEO is released at the aggregate level of U.S. census division of which there are nine in the country (Figure 1), and energy use is divided between four demand sectors: commercial, industrial, residential, and transportation. This project will focus on the residential and commercial sectors only.

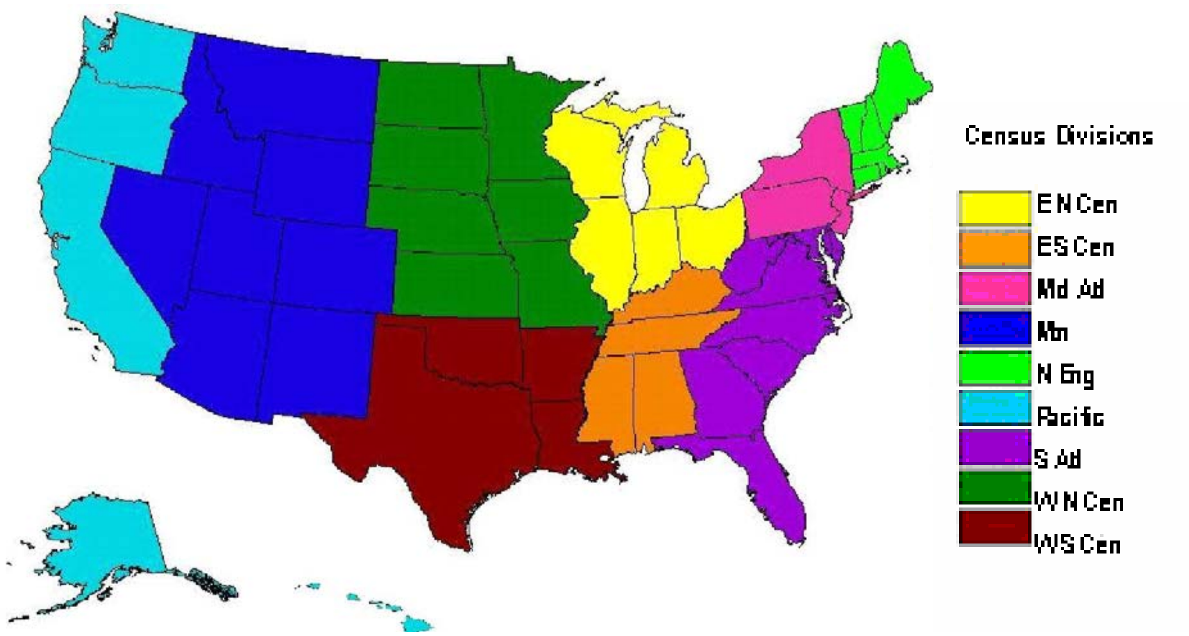


Figure 1: The nine U.S. census divisions

1.1 Study Area

The U.S. geography is characterized by a highly skewed distribution of population and energy use with the preponderance of activity on both coasts. It is important to note that Alaska and Hawaii have not been added to the following maps in this paper, but were a part of the analysis. They were omitted because of their distance from the continental U.S. and their lack of influence on the output results.

1.2 Energy Analysis

1.2.1 NEMS

NEMS contains 13 modules (Figure 2) that model the interactions of energy markets and provide insights into future changes in supply, demand, economic conditions, etc. NEMS outputs are used to analyze economic policies, technological changes, changes in legislation, and other energy related topics. Each module of NEMS contains many assumptions that characterize the future production, conversion, or consumption of energy in the U.S.

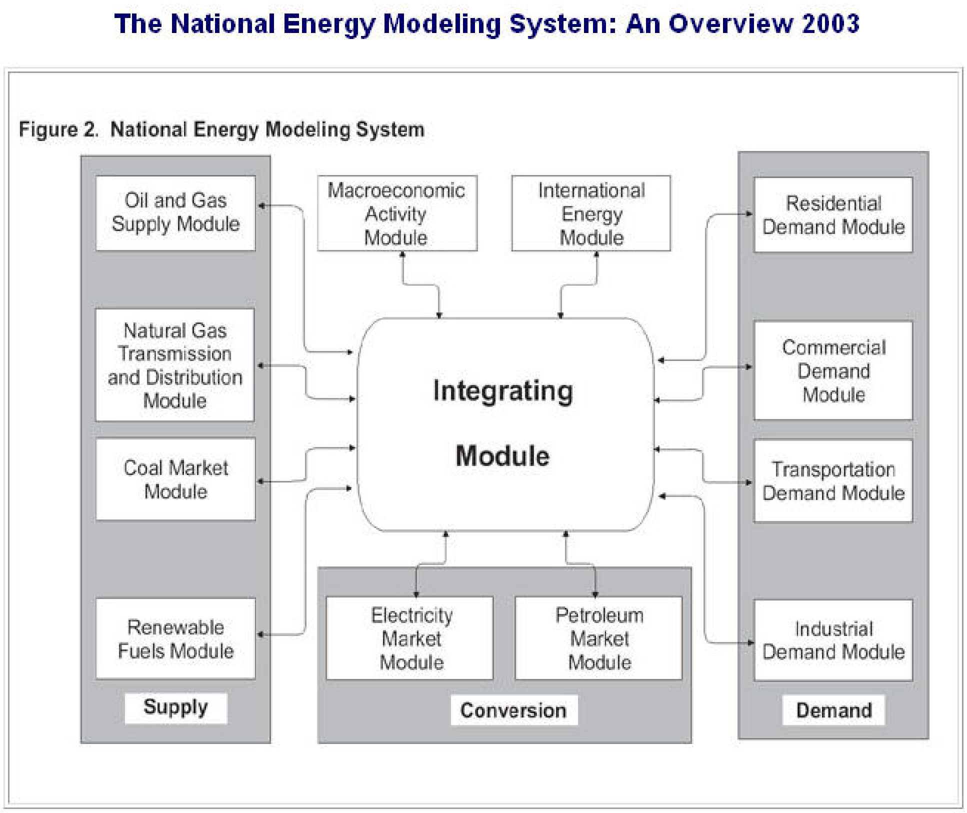


Figure 2: The 13 modules of NEMS

source: (Energy Information Administration 2003b)

The AEO provides a number of alternative forecast scenarios derived using NEMS, including variation in world oil prices and 21 other cases that explore the impacts of varying key assumptions in the individual components of NEMS. Many of these cases involve changes in the assumptions that impact the penetration of new or improved technologies, which is a major uncertainty in formulating projections of future energy markets.

1.2.2 NEMS Electricity Market Module

The NEMS Electricity Market Module (EMM) is one of the many modules within the NEMS framework. During a model year, the EMM receives three main categories of inputs (for details of inputs see Figure 3):

- Electricity demand from the NEMS demand modules
- Fuel prices from the NEMS fuel supply modules
- Macroeconomic parameters from the NEMS macroeconomic module

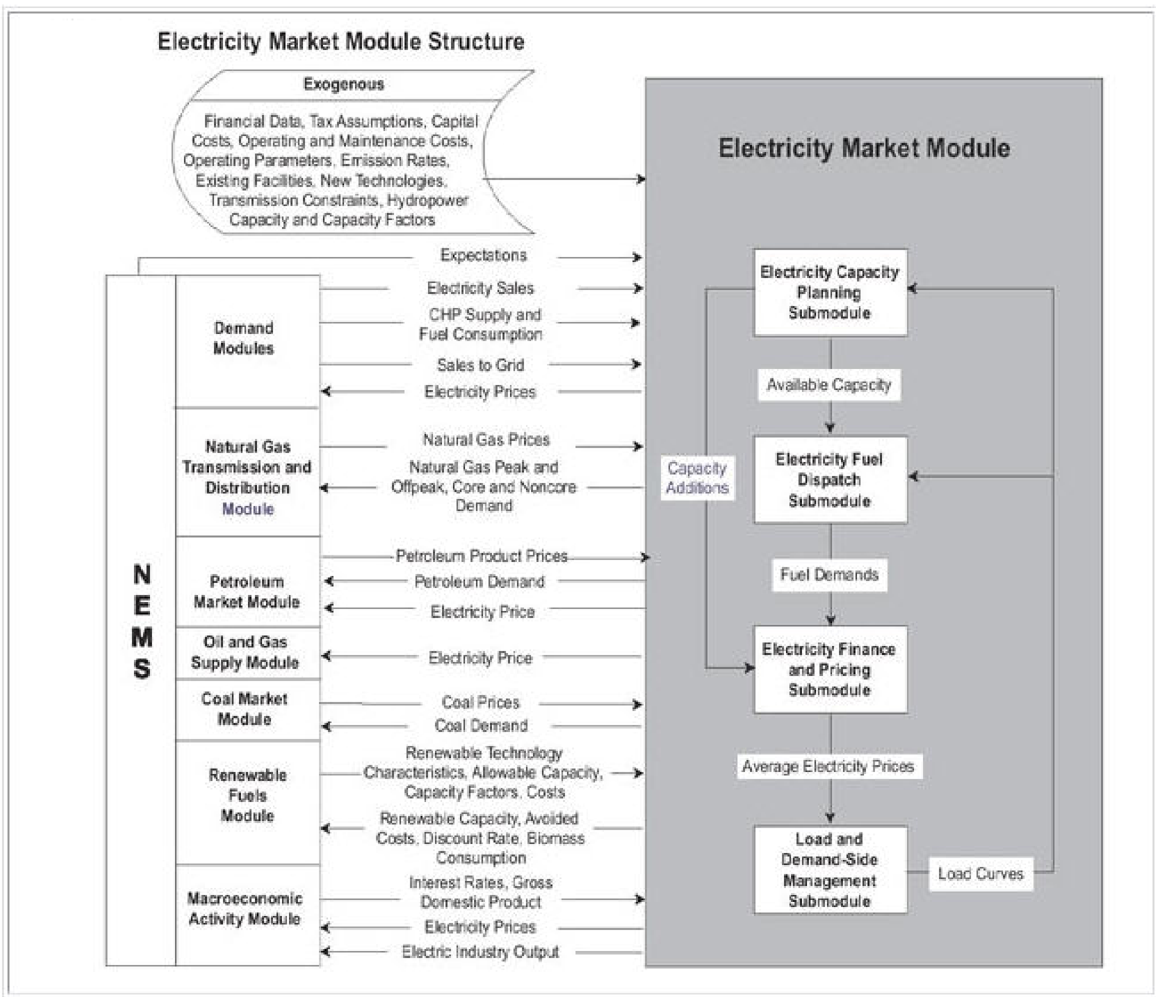


Figure 3: Detailed list of input and outputs for the Electricity Market Module

source: (Energy Information Administration 2003b)

With these inputs, the EMM forecasts the actions taken by electric utilities and non-utilities (off-grid electricity generation) to meet demand in the most economical manner. This forecast of electricity supply is the primary data source for the current project.

1.3 Problem Statement

A part of LBNL's research encompasses the integration of small-scale regional distributed generation facilities into NEMS. This type of analysis creates a demand for national level data in a more granular, higher resolution form. Clearly, there is a need to create methods to spatially disaggregate the NEMS model's intermediate results and outputs in a consistent, repeatable, and defensible manner. Since the NEMS data are generated on an annual basis for every year of the forecast period, it is important to have this method automated for both consistency and expediency. The value of the final product lies in the fact that it maintains a national scope while providing data at the local level, and will be useful to those analysts looking at national energy trends on a regional scale. And more importantly, results should be consistent with the structure and latest assumptions of NEMS and so can be reported as adjuncts to AEO results fed back into NEMS to enhance other analyses.

1.4 GIS Solution

Disaggregation is the process of separating an aggregate body into its component parts. The immediate goal of this project is to determine the appropriate method of disaggregation for the NEMS electricity forecast data that will be applicable to the analysis of the commercial and residential end-use sectors. This approach will be demonstrated by choosing one end-use sector (residential energy) and implementing the methods and procedures to disaggregate NEMS outputs to the county level. The flexibility of this procedure will allow it, in the future, to be expanded to other end-use sectors and to energy carriers other than electricity, such as natural gas or hydrogen.

1.5 Scope and Methodology

This effort has been executed as a pilot project to demonstrate the potential benefits of GIS to improve the effectiveness of NEMS analysis. This work provides insight into methods available for spatial disaggregation, adequacy of the method implemented, and procedures needed to automate the method.

The method employed for this project can be summarized in the following steps:

1. Problem Identification – The problem areas in the current NEMS data workflow were identified.
2. Literature Review – Research was undertaken into methods of spatial disaggregation and their applicability.
3. Interface Customization – The ArcGIS interface was customized to perform specific tasks required for the NEMS disaggregation process.
4. Time-Series Animation – The disaggregation outputs of 25 years of electricity demand data were visualized in a time-series animation.

2. Methods for Spatial Disaggregation

Disaggregation is the process of transferring attributes from one set of spatial objects to another within a defined portion of geographic space (see Figure 4) (Goodchild et al. 1993). There are a number of methods available to assist an analyst to disaggregate data into new areal configurations.[1] The approaches can be grouped into two categories: vector- and raster-based methods. Vector disaggregation methods, called *Areal Interpolation*, rely on an alternative often non-nesting set of areal units to take data from one unit and distribute it within a differently-sized unit (see Figure 4). The raster methods require the use of surface modeling which bases analysis on a continuous surface of population data. Below the two methods are detailed and compared, leading to selection of the method most appropriate for the project.

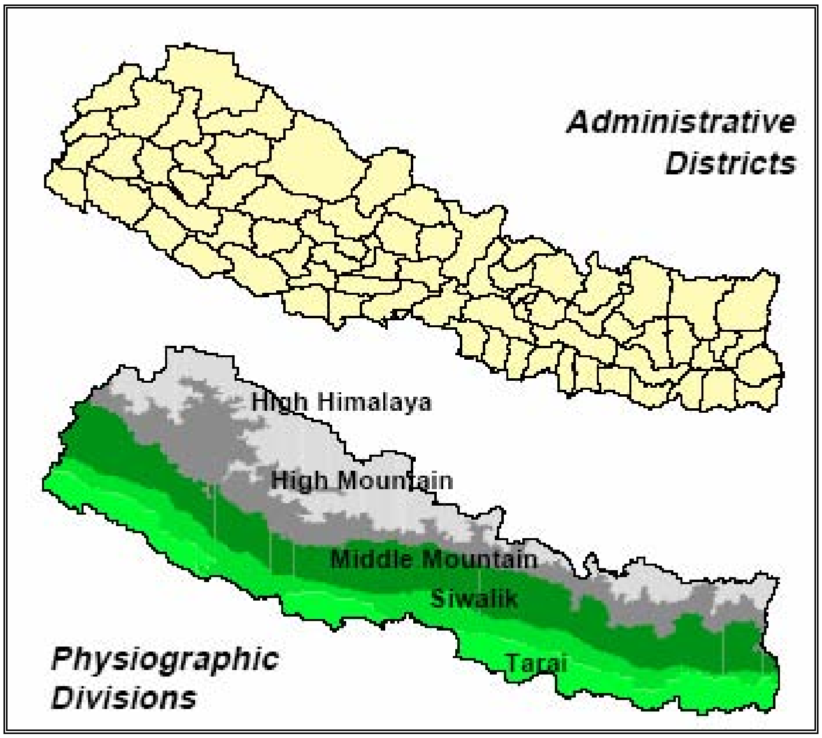


Figure 4: The process of Areal Interpolation.

source: (Deichmann 1996)

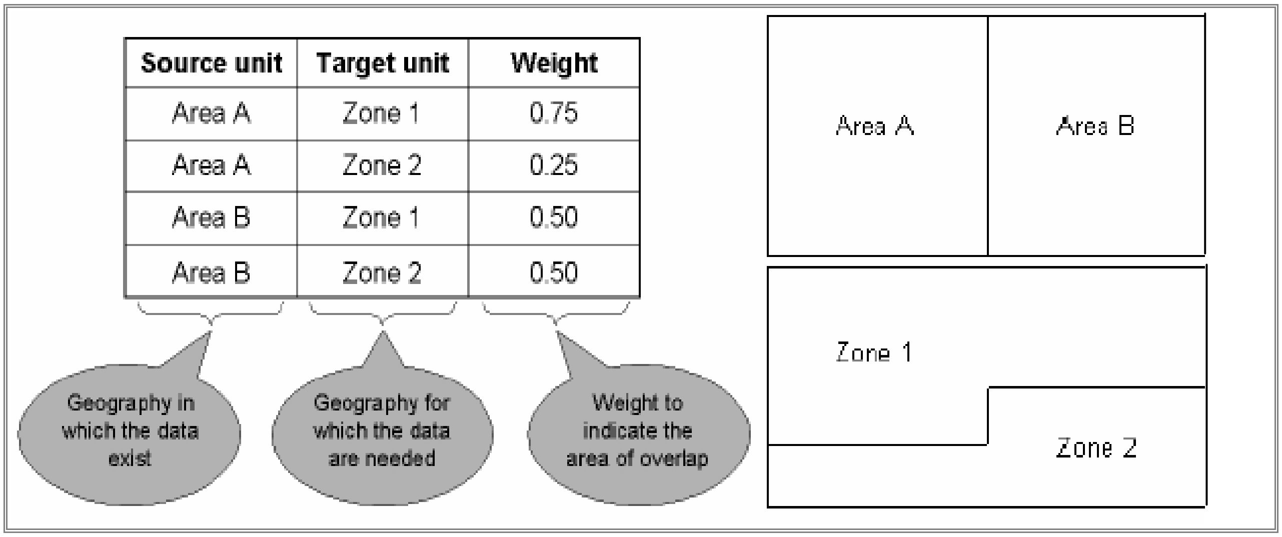
2.1 Vector-based Methods

After Deichmann (1996), the areal units of the input data are called "*Source Zones*", the output areal units are the "*Target Zones*" and the areal units of ancillary data are the "*Control Zones*". The following sections below describe two approaches to vector-based disaggregation.

[1] The term *areal* describes something that relates to or involves an area.

2.1.1 Areal Weighting

In the case of areal weighting, all of the Source Zones are homogeneous in value distribution (Figure 5). The target zones (P_t) are simply overlaid on the source zones (P_s) and the values are allocated according to the proportion (A_{st}) of each source zone falling in each target zone (see Equation 1). The obvious problem with this method is that few phenomenona are distributed through space homogeneously (Norman and Marsh 2003).



Source unit	Target unit	Weight
Area A	Zone 1	0.75
Area A	Zone 2	0.25
Area B	Zone 1	0.50
Area B	Zone 2	0.50

Figure 5: An example of the Areal weighting technique

source: (Norman and Marsh 2003)

$$\boxed{P_t = \sum_s P_s \left(A_{st} / \sum_t A_{st} \right)}$$

Equation 1: The Areal weighting calculation

where:

P_t = Population Target Zone
P_s = Population Source Zone
A_{st} = Area of overlap between source and target zones

2.1.2 Areal Weighting using Control Zones

In the case of areal weighting using control zones, ancillary data are incorporated into simple areal weighting to help improve target zone estimates. For example, control zones might be the areas of land and water that would be used to mask out the areas within the target zones that are not inhabited. The control zones boundaries are not required to match either the source or target zones, which allows for flexibility in choosing a suitable dataset. The control zones can be assigned a density or weight through a mathematical or heuristic method (Figure 6). The control zone density or weight, d_c, can be applied to the calculation with the overlap between the control and target zones, A_{ct}, to solve for the population (or another count variable) of the target zone, P_t, as seen in Equation 2 (Goodchild et al. 1993).

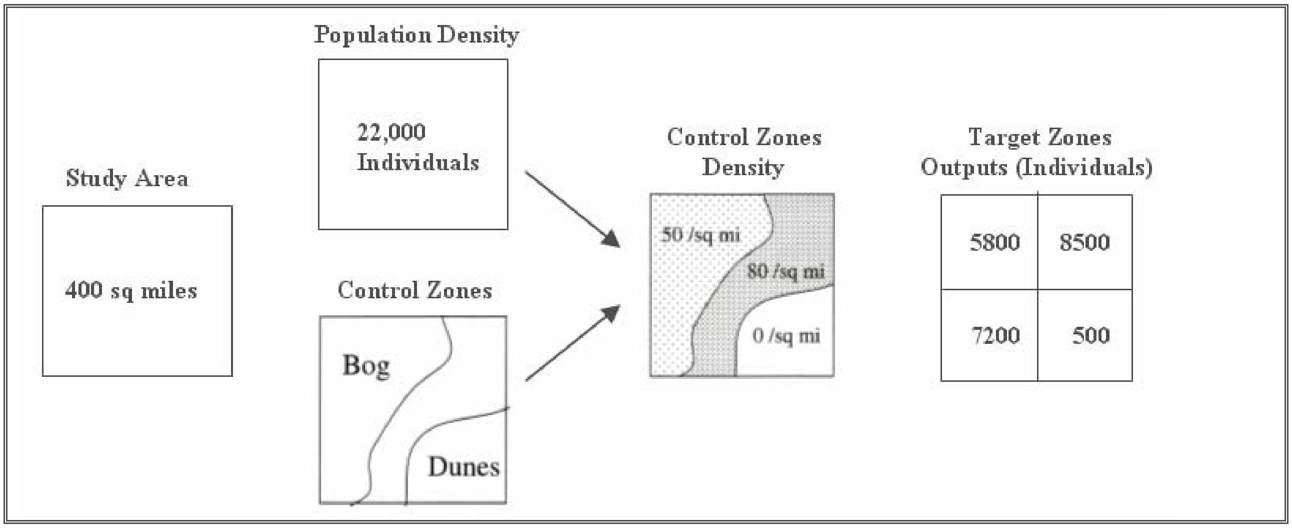


Figure 6: An example of the Control Zone technique

$$\boxed{P_t = \sum_c d_c * A_{ct}}$$

Equation 2: The Areal Interpolation with control zone calculation

if ($n_c < n_s$)

where:

n_c = number of Control Zones
n_s = number of Source Zones
d_c = Density of Control Zone
P_t = Population Target Zone
A_{ct} = Area of overlap between Control and Target zones

It must be noted that with both of the above methods, there is a large chance of error due to data coarseness. Clearly, it is always advisable to obtain higher resolution data when available.

2.2 Raster-Based Methods

Raster-based methods produce a regular grid with each cell containing an estimate of the total count value representative of the particular location. There are certain advantages to these methods. For example, there is the ease of reaggregating the data into any size areal unit necessary. Another advantage is compatibility of the data with the large amount of other information available in raster format, such as remotely sensed data, environmental data, etc. Using these methods also retains, within each of the grid cells, the underlying data of the aggregated phenomenon so that it is possible to reconstitute the original data if the resolution is small enough.

2.2.1 Centroid-based Distribution

The centroid-based distribution method uses polygon centroids attributed with the appropriate data values to represent the source zones. The values can be distributed over a raster by using a kernel function, which takes into account the local density of centroids with a suitable distance decay function (Figure 7). This allows for discontinuities within the output grid. The result of this process is an output grid containing cells equivalent to the attribute value for that cell's area (Deichmann 1996).

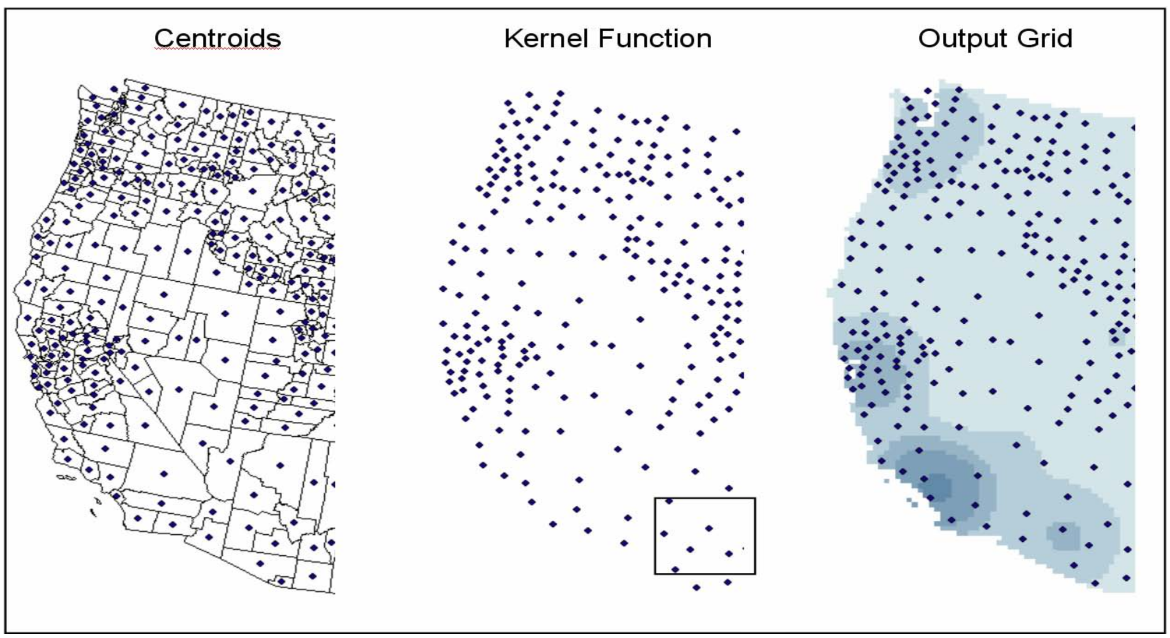


Figure 7: The centroid-based distribution technique

source: (Deichmann 1996)

2.2.2 Maximally Smoothed Interpolation

The maximally smoothed interpoloation method, also called "*Pycnophylactic Smoothing*" (Tobler et al. 1997), requires a grid with discrete areal units. The motivation behind Pycnophylactic Smoothing is preservation of the source zone area while varying the internal data density. The technique is an iterative one (Figure 8), where the distribution within an area is slowly reallocated using the rationale "things are more similar in nearby locations". This allows cells near the boundary to migrate towards the value of adjacent areas while still keeping the total value of each source zone. This technique is recommended only if there is little or no additional information available and if the units are relatively homogeneous (Deichmann 1996).

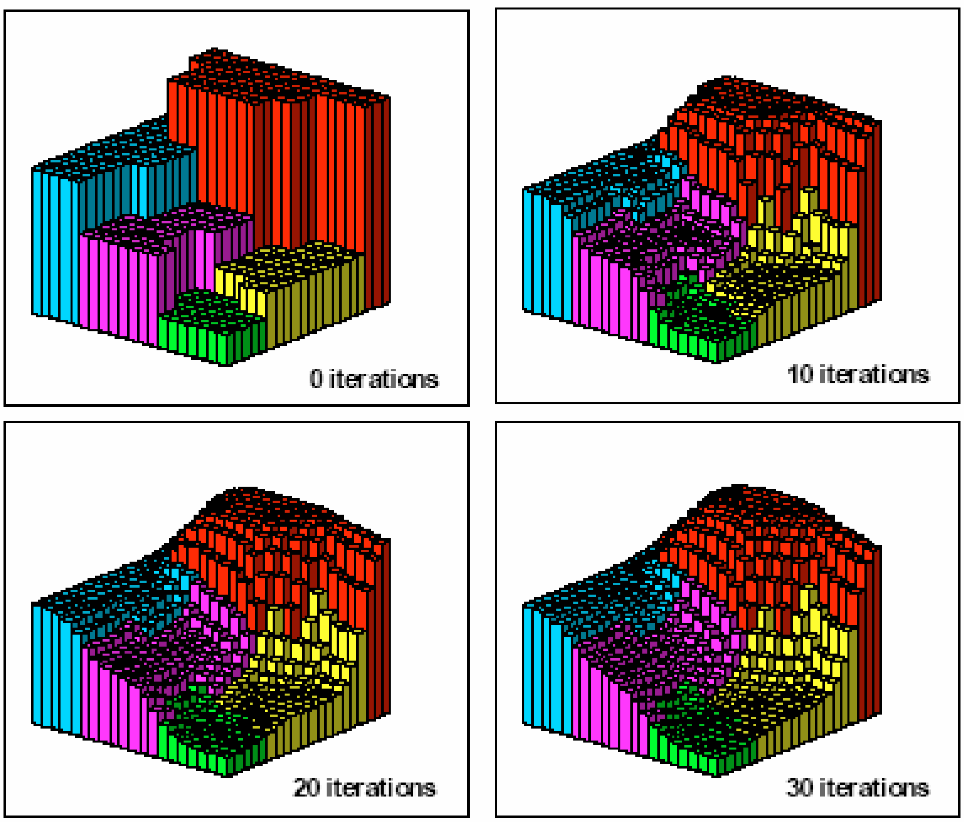


Figure 8: The iterative area conserving process of Pycnophylactic Smoothing

source: (Deichmann 1996)

2.2.3 Smart Interpolation

Smart interpolation is based on heuristic knowledge of a subject and the availability of related data sets (Figure 9). *Smart Interpolation* uses secondary data as influence factors, as one might in a typical grid cost-weighted analysis scenario. The first step creates a mask in order to eliminate all of the regions where the phenomenon is not pertinent. The second step involves creating an *influence surface* composed of the ancillary data. Each influence surface is heuristically adjusted to assign a relative weight to each cell in the grid. Finally, the source zone values are distributed to the corresponding target grid cells in proportion to the weights constructed in the previous steps. The strength of this method is that human phenomenon are not distributed randomly across space but are related to geographies with certain characteristics (Goodchild et al. 1993).

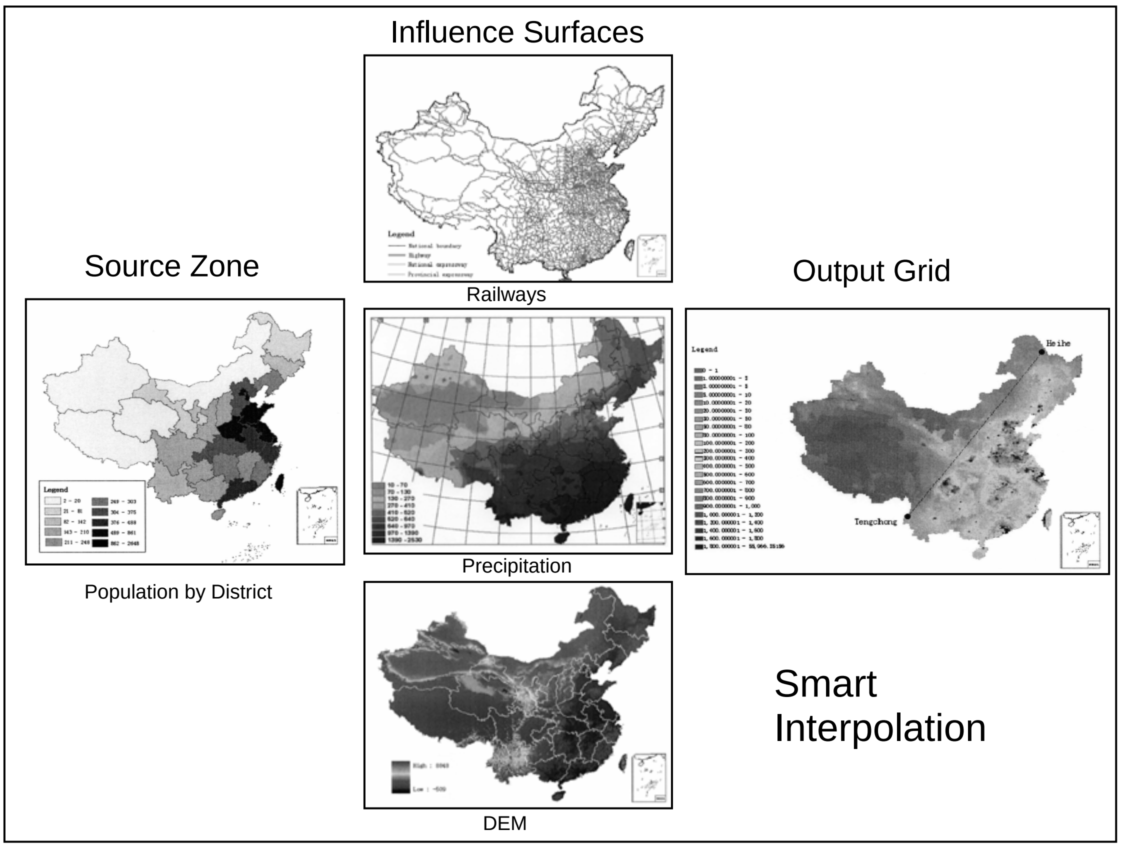


Figure 9: A Smart Interpolation map of China

source: (Yue et al. 2003)

Figure 10 clearly shows a higher level of detail produced by the Smart Interpolation method used to distribute population across Africa (Deichmann 1996). The question to be asked is whether this level of detail is an accurate representation of the phenomenon. More light will be shed on this question in Chapter 4.

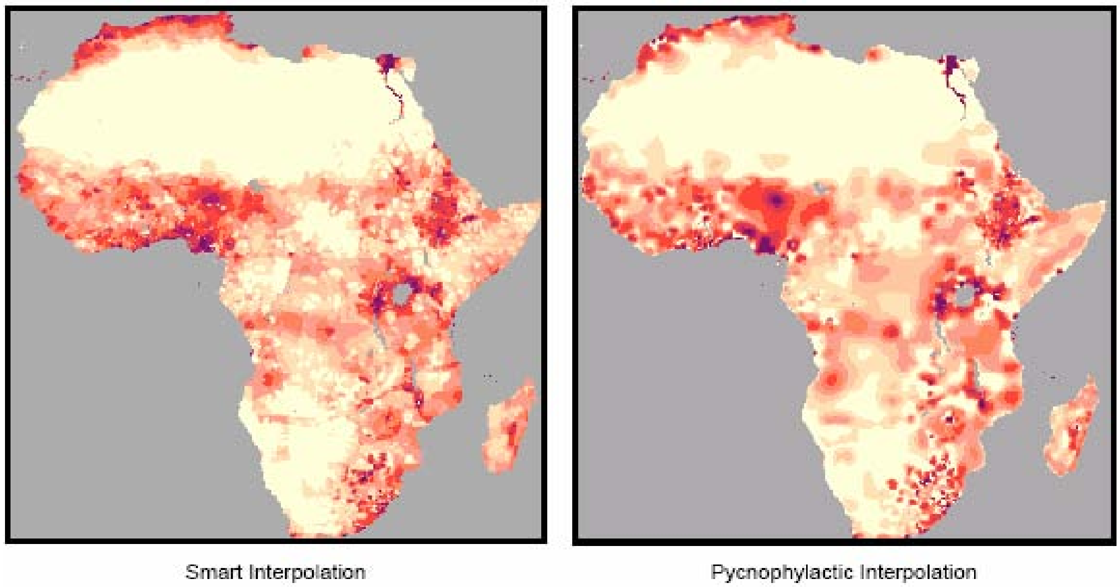


Figure 10: A comparison of two raster-based Areal Interpolation methods (Africa)

source: (Deichmann 1996)

2.2.4 Cartographic Approach

Cartographic approach, as shown in Figure 11, is seldom used and is primarily for the visualization of spatial data as opposed to the analysis of data. Since there is no analytical backing for this approach it will not be discussed further. For additional information see (Leddy 1994).

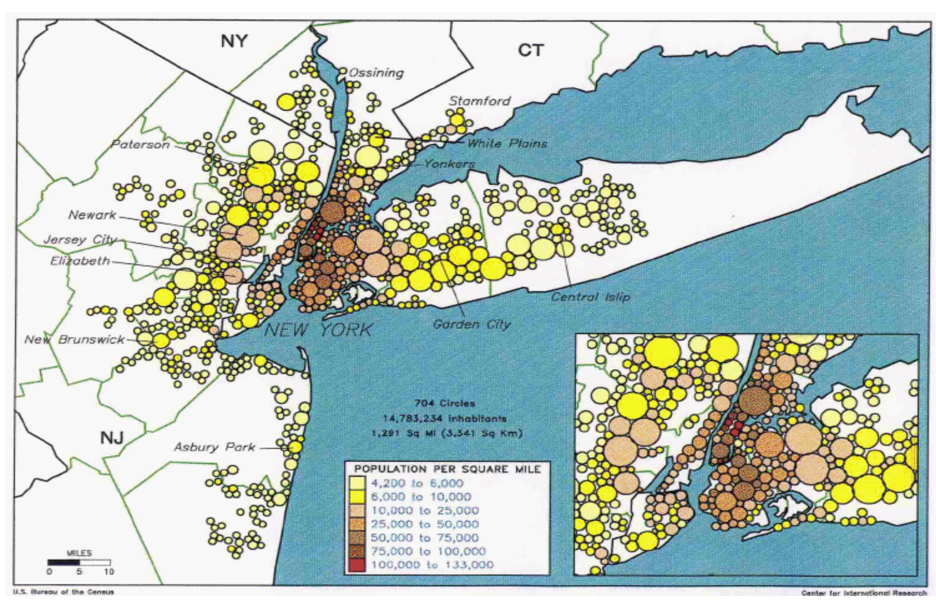


Figure 11: The cartographic approach used for visualization

source: (Leddy 1994), http://www.ciesin.org/datasets/cir/gifs.gpopdb.html

2.3 Project Methodology

The overall approach of this project is to investigate the correlation between weather and energy demand. To be consistent with this objective, weather data will be integrated as a control factor into the analysis of country-wide energy demand. LBNL works primarily in vector format for two reasons. First, the majority of energy data is compiled in vector format as it is often derived from the facilities management field. Second, much of the data is supplied as tables associated with geographic regions, a logical step from vector analysis.

Areal weighting using control zones was the technique used based upon reviews of methods for spatial disaggregation. The main reasons for this choice, outlined in the remainder of the chapter are application needs and preferences, available expertise and feasibility. The final section is devoted to describing the details of the methods involving the control zones and the key aspects of the functional model.

2.3.1 Feasibility

The feasibility of the disaggregation method becomes a larger issue when looking at project length and the learning curve inherent in certain technologies. One of the most important considerations for this specific case is the ease with which the process could be coded into an application. The vector objects within the ArcObjects framework are significantly easier than raster objects to manipulate within the ArcGIS development environment. Another consideration is the file size. Since the study area is large in scope, raster files of the necessary resolution would also be quite large and would require long and expensive computation. The last major consideration is the ease in which the initial data values are preserved within the data structure given that the counties (Target) nest with the census regions (Source). Although preservation of data is possible in the raster structure, there are more steps necessary to obtain similar results. Using a vector format, the actual values of electricity demand can be preserved in each stage of the process. In comparison, the data values of a raster framework must be recalculated after being integrated into the structure.

2.3.2 Control Layer and Weighting

The Areal Interpolation with Control Zone method requires source, target, and control zone data to assist in the reallocation of aggregated data (Figure 12). In this case, climate zones, discussed in the Energy Demand Data section (Section 3.4) will act as the control zone. Using climate zones relies on the assumption that the majority of electricity use is in warmer climates primarily due to air conditioners but also appliances such as fans and refrigerators. Another effect is on the extremely cold regions where people tend to spend more time inside and there are longer nights. There is also the assumption that residential electricity intensity will differ from commercial electricity intensity in each of the climate zones.

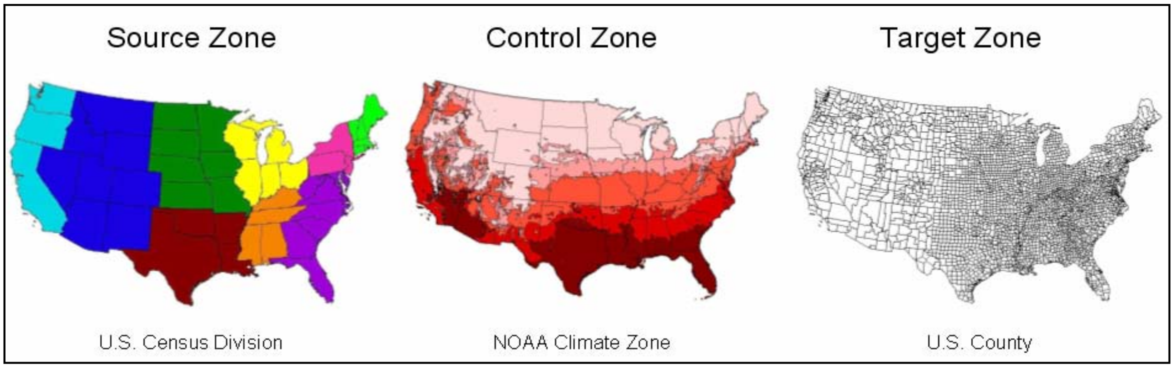


Figure 12: The three areal zones used in the analysis

Weighting values for the five climate zones taken from the *Commercial Building Energy Consumption Survey* (CBECS) were used in the areal weighting calculations (Energy Information Administration 2002). The commercial intensity values for each climate zone are derived from electricity use per worker, and residential intensity values are derived from energy use per household. These values act as weighting factors to distribute the electricity use to the warmer regions of the country. The Commercial and Residential weighting values are, respectively, from the 1999 Commercial Building Energy Consumption Survey and the 2001 Residential Energy Consumption Survey (Energy Information Administration 2003a). Below are the weighting factors to be used in this calculation (Table 1).

Control Zones	Commercial Weighting Factor	Residential Weighting Factor
Zone 1	11.0	9.6
Zone 2	10.7	8.99
Zone 3	10.5	10.3
Zone 4	11.5	10.1
Zone 5	12.0	13.6

Table 1: The weighting factors of the control zones

2.3.3 Functional Model

The functional model is shown in Figure 13 as a series of geoprocessing and tabular steps that successfully disaggregate the NEMS outputs. The model is considered deterministic in nature and provides hard numbers as output, eventually providing each county with a proportion of the entire census region's electricity demand. This model does not allow for fuzzy functions or probability, a possible future enhancement.

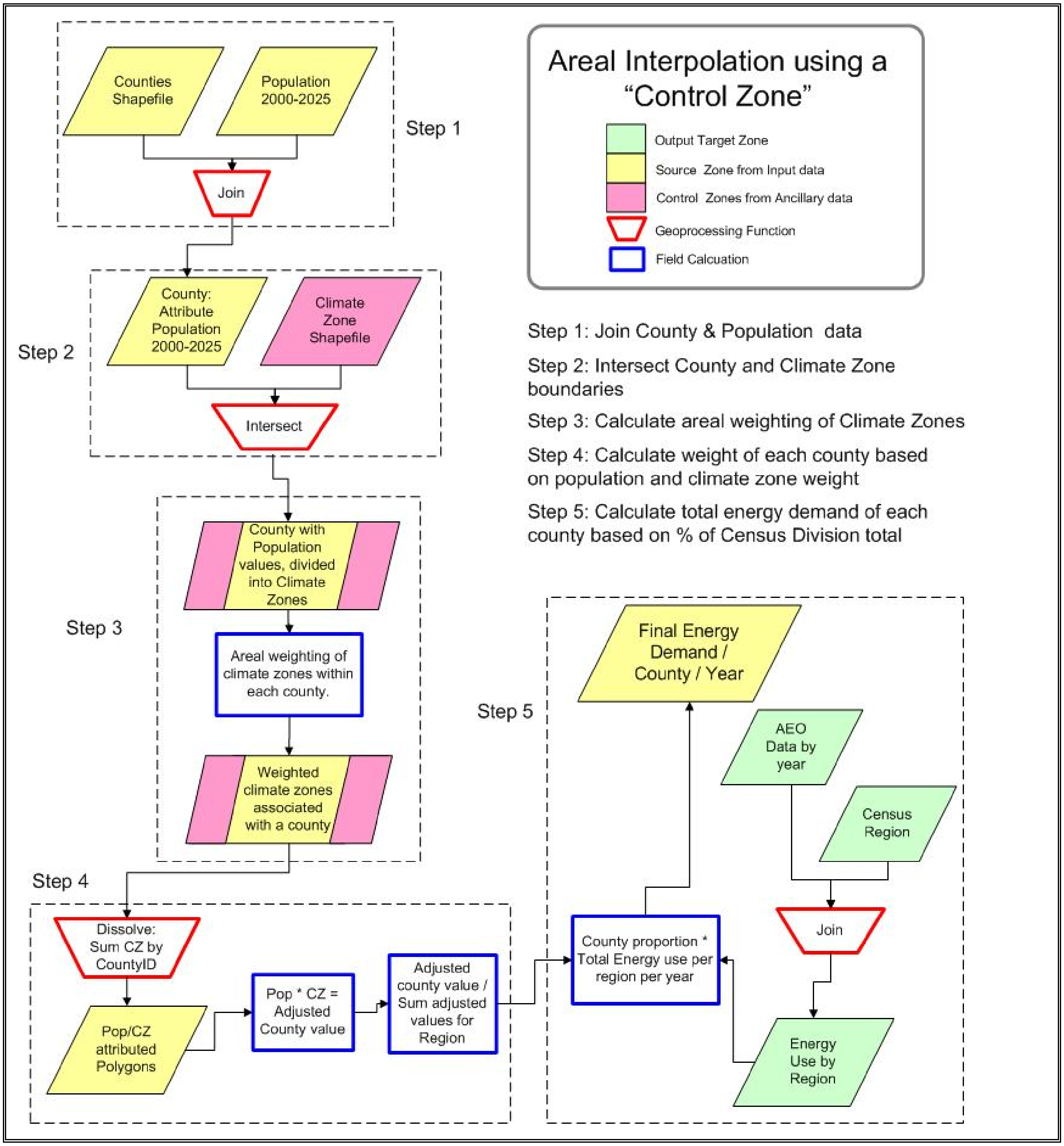


Figure 13: The functional model used to disaggregate the NEMS outputs to the county level.

Figure 13 shows the process broken down into five steps using San Diego County as an example; these steps utilize both geoprocessing and tabular functions.

1. Join County border shapefile and population table to assign population values to each county.

2. Intersect County border shapefile and climate zone shapefile to create polygons of each climate zone within each county.
3. Calculate the value of climate zones within each county using areal weighting. After the intersect (step 2), each county will encompass one or more polygons from different climate zone values. The area of each new polygon is calculated and divided by the total area of the original county to get the proportion of the total area. This proportion is multiplied by the associated climate weight. All polygons for each county now have a climate zone influence value based on their proportional area of the county.
4. Calculate the weight of each county in a census division based on their population and climate zone rating: The dissolve function is then used to reassemble each county into one polygon. The climate zone influence values are summed up to form a composite weight for the entire county. Each county now contains population values (2000-2025) and a single climate zone weight. The two values are multiplied to form a value that indicates the population/climate influence of the county within the census division.
5. Calculate the total electricity demand of each county based on their percentage of the Census division total: The population/climate influence values for each county are summed for the census division. Each county is then divided by this sum to create a proportional influence on the census division. The proportion for all the counties within each Division will always sum to one. The proportion influence for each county is then multiplied by the census division Energy Demand total for the year in question. Each county will be allotted their proportion of the total, and all the counties energy demand values within each census division will always sum to the total for the census division.

Figure 14 shows graphically the same process as performed with the GIS software program.

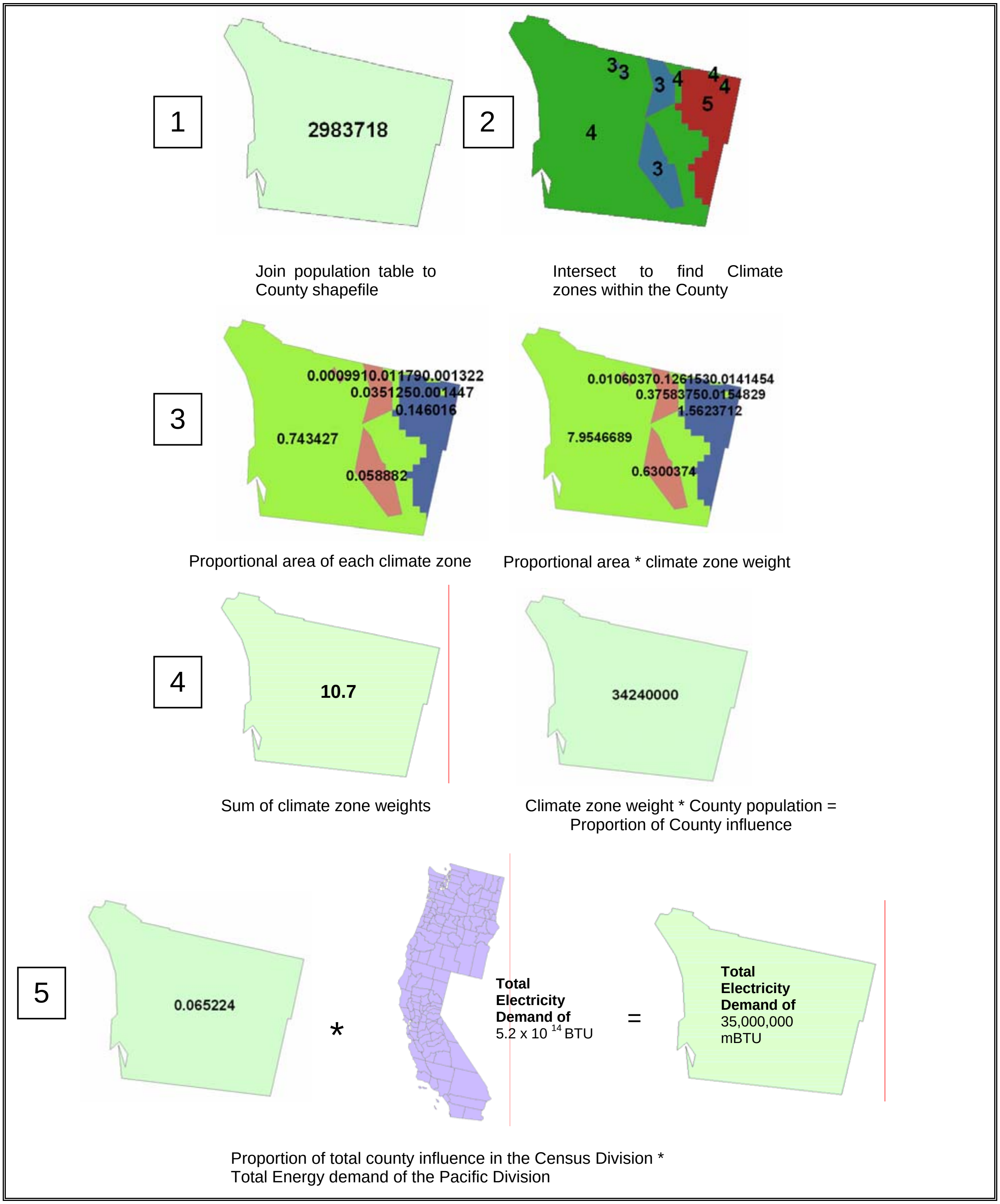


Figure 14: A step-by-step example of the disaggregation model performed on San Diego county, CA.

3. Data

The data for disaggregation, obtained primarily from web sources, were closely linked to the chosen method. The data set was projected in Albers Equal Area, NAD 83. At the outset there was a list of potential data that included a complex variety of influence factors, but this list was narrowed down by needs assessment. The data required intensive statistical preprocessing using Excel and Access, as well as spatial preprocessing using ArcGIS.

This chapter details the preprocessing steps completed on each dataset as well as data strengths, weaknesses, assumptions, and overall validity. The steps included:

- Statistical completion of datasets
- Create naming convention
- Check topology
- Survey completeness
- Name matching
- Clipping
- Reprojection/transformation
- Metadata creation

3.1 Overview of Datasets

The project required the following data for input into the model (datasets 1-4) and testing (datasets 5-6).

1. **Population Data**: population forecasts for all counties in the USA for the years 2000-2025 (U.S. Environmental Protection Agency 2004)
   Issues: incomplete, unknown forecasting methods used

2. **County Borders**: all counties in the USA (U.S. Census Bureau 1997)
   Issues: multi-polygon counties

3. **Electricity Demand Data**: Electricity Demand Forecasts by census division for every fifth year during 2000-2025 (Energy Information Administration 2004a)
   Issues: complex input assumptions

4. **Climate Zone Data**: series of five climate zones derived from the annual heating degree day and cooling degree day maps (U.S. Census Bureau 1997)

5. **Comparative Testing Data 2000**: the official electricity use data by state, broken down by end-use (Energy Information Administration 2004d)

6. **2000-2012 Energy Outlook Data:** collection of electricity use statistics by county (California Energy Commission 2002)

3.2 Population Data

The Environmental Protection Agency (EPA) maintains a website containing population forecast datasets (U.S. Environmental Protection Agency 2004). It provides annual totals for each year from 2000 to 2025. There are no statutory requirements for population forecasting; therefore each county determines its own need for this type of data. The population data acquired for this project were compiled by an EPA analyst who contacted each and every county or state to obtain their most recent population forecast. Because of this, there are counties with forecasts ranging from 0 to 30 years. There were two states, South Dakota and Iowa, that had no forecasts. This issue was handled by extrapolating a linear growth rate from past census years.

Figure 15 shows the distribution of population across the contiguous states (U.S. Environmental Protection Agency 2004). The large counties in the west have similar population totals as the smaller counties in the east, but represent entirely different densities and energy demand. Wide variations are found across the country when population is plotted as a function of county area. Thus it is customary to normalize population data, often by area, to produce a density map. A perfect example of this phenomenon is the comparison of the population density map (Figure 16) which shows a higher than overall density east of the Mississippi river, and the population count map (Figure 15) which only shows higher values around the big cities and the large counties in the southwest.

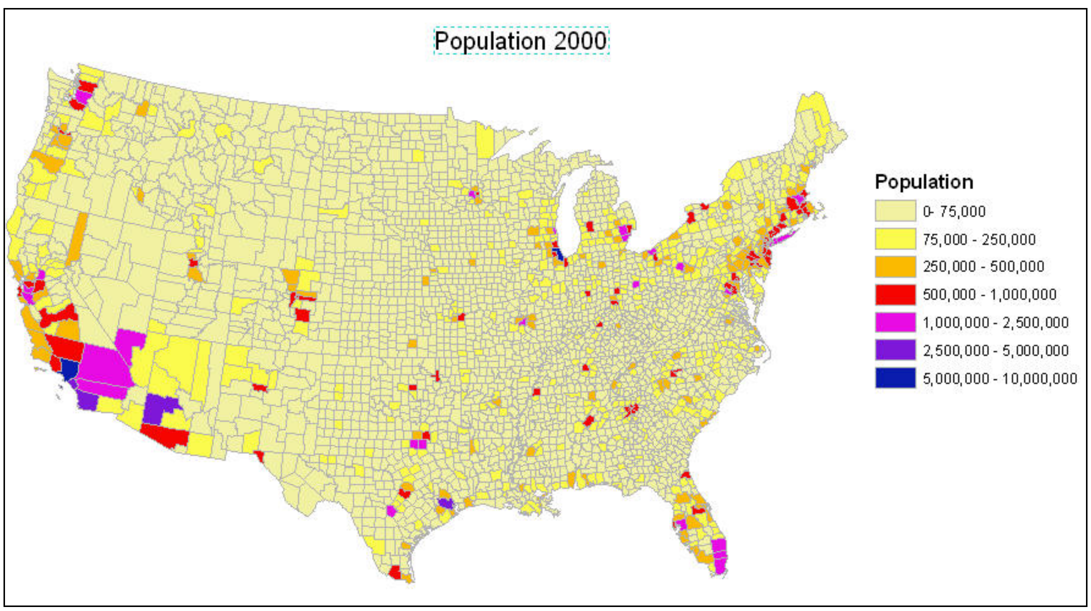


Figure 15: Population counts by county for 2000

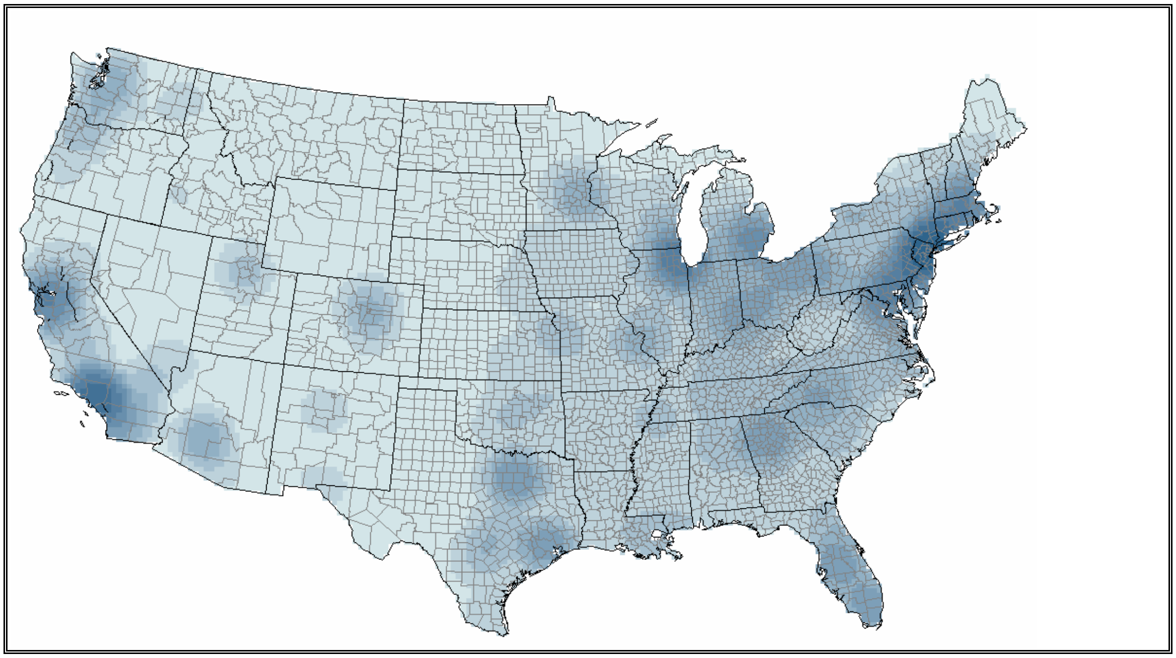


Figure 16: Population density ranging from high density (dark blue) to low density (light blue) for the year 2000

3.2.1 Statistical Preprocessing

The first step, after acquiring the EPA population data, was to separate the counties into the prescribed census divisions to help locate missing information. Figure 17 is a snapshot of the 25 years of the raw forecasted data for a handful of counties and the missing information ("gaps") that existed prior to processing.

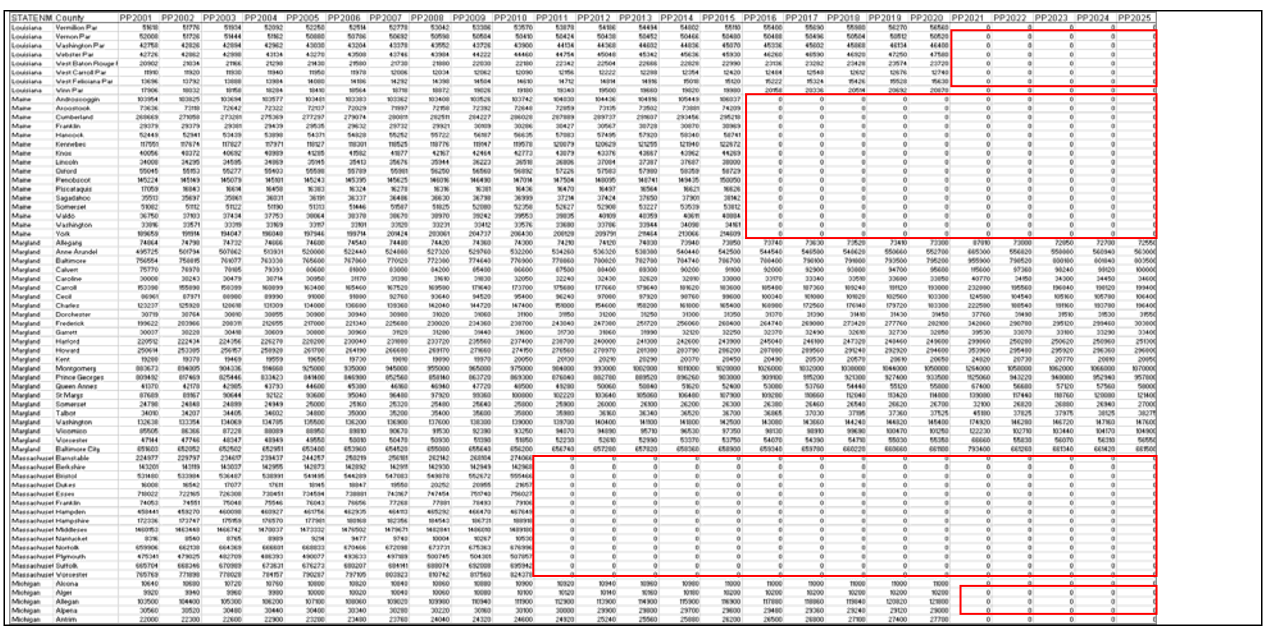


Figure 17: An example of the "gaps" of missing information (rectangles) in the population spreadsheet.

Due to the varying extent of missing data (Table 2), two statistical methods were used to deal with this discrepancy in the data set as illustrated in Figure 18.

Population Model Completeness Survey

No Data Processing		Minimal Data Processing		Major Data Processing	
Alabama	2025	Connecticut	2020	Alaska	0
Arizona	2025	Delaware	2020	Arkansas	2012
California	2032	Florida	2020	Georgia	2010
Colorado	2025	Illinois	2020	Hawaii	0
Idaho	2025	Indiana	2020	Iowa	0
Kansas	2030	Louisiana	2020	Maine	2015
Kentucky	2025	Michigan	2020	Massachusetts	2010
Maryland	2025	Nebraska	2020	Mississippi	2015
Minnesota	2025	Nevada	2022	New Jersey	2015
Missouri	2025	North Dakota	2020	Ohio	2015
Montana	2025	Pennsylvania	2020	South Dakota	0
New Hampshire	2025	Rhode Island	2020	Vermont	2015
New Mexico	2025	Tennessee	2020	Virginia	2010
New York	2025	Wisconsin	2020	Wyoming	2010
North Carolina	2025				
Oklahoma	2025				
Oregon	2025				
South Carolina	2025				
Texas	2025				
Utah	2025				
Washington	2025				
West Virginia	2025				

Table 2: Completeness survey of forecasted population data

The first method of preprocessing was a simple linear function interpolating from the available data (Table 3). This is the standard method used to interpolate missing data. The two states with no county projection information were South Dakota and Iowa. This process was applied to the only available data were population totals by county for the last two census counts (1990 & 2000). These figures were used to create a linear growth rate function projected out to 2025.

Population	1990	2000	2010
County A	10,000	12,000	= 12,000* 1.2
			= 14400
Growth Rate	= (2000/1990)		
	1.2		

Table 3: Calculating a linear growth rate for counties with no population forecast

A common criticism of population forecasting is the assumption of linear growth. Almost all forecasts will use a linear function to describe the growth of a population, which in reality is not possible. Although in theory it might be advantageous to find a more sophisticated method of forecasting population, the reality is that most techniques lose their value when projecting 25 years into the future. Still, it must be noted that the validity of the population forecast data is questionable. Fortunately, these two states have very little bearing on the energy demand and energy usage in their respective census divisions, so their influence on the final outcome is minimal.

The second method is a more complex technique that utilizes the past ten years of county population data and forecasts of the region population model to calculate a linear time series model. This method takes advantage of the already forecasted data and the region totals, allowing for variation over time. The majority of the "gaps" in the county population dataset were filled in this manner.

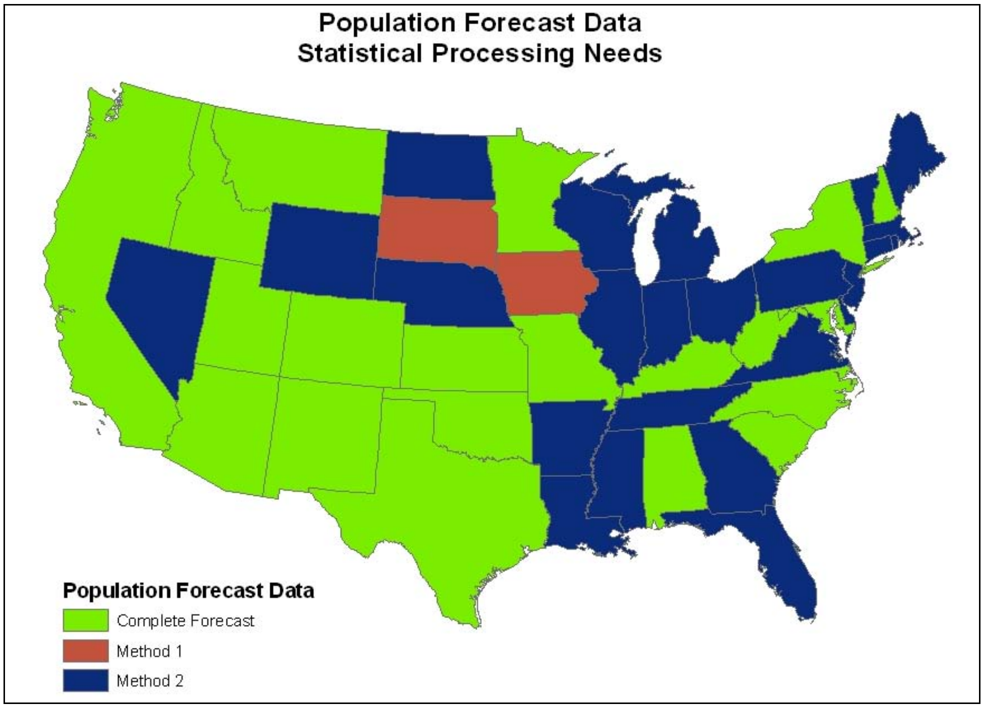


Figure 18: Different methods used for forecasting population data by state

Equation 3 takes the sum of the previous ten years of county projections (*PopCounty*) and divides that by the sum of the previous ten years of the region's projections (*PopRegion*). This number gives a proportion that is then multiplied by the region's (*PopRegion*) current year (11) to get the final prediction for the county projection for year 11.

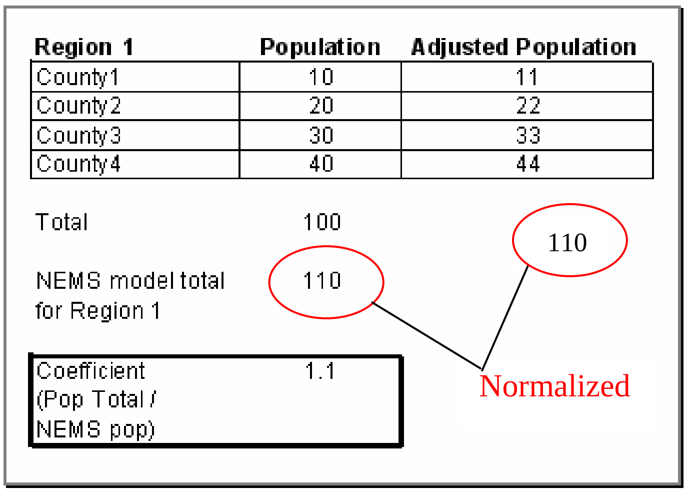
$$PopCounty_{11} = \left(\sum_{i=1}^{10} PopCounty / \sum_{i=1}^{10} PopREGION\right) * PopREGION_{11}$$

Equation 3: The primary equation used to forecast missing population data

3.2.2 Normalizing Population Data

NEMS uses certain population assumptions for each census division for each forecasted year. Unfortunately, the EPA population forecast by county does not sum up to the same census region totals. To deal with the differences, the following process was conceived. To integrate the two population forecasts a baseline population for each census region in each relevant year was required, and the NEMS population assumption was used.

The EPA population forecast totals aggregated to the census region was subtracted from the NEMS total, and that difference was divided by the NEMS census region total (Figure 19). This outputs a coefficient that can be multiplied by all the county values for the year at hand, hence making both the calculated and NEMS population totals equal.



Region 1	Population	Adjusted Population
County1	10	11
County2	20	22
County3	30	33
County4	40	44

Figure 19: Steps to normalizing the population data

3.2.3 Forecasting Issues

There is a wide variety of opinions on the validity and usefulness of population forecasting models. Because there are so many factors influencing each region and country in the world, political, economic and environmental, it is nearly impossible to forecast more than 5 years into the future.

> “the reliability of forecast data depends on two major assumptions…. First, that the relationship between whatever underlying factors and the overall growth rate are perfectly understood….. and second that all of the factors contributing to an overall growth rate will occur as planned. The forecast is only telling you the likely populations if the assumed conditions and causes should actually occur” (Zepp 2004)

Another author speaks to the deterministic point of view in population modeling:

> ”In other words, one assumes perfect correlation between two variables such as fertility and mortality…. Assumptions of this kind are unrealistic, and, moreover,

> they cause inconsistencies: two variants that are extreme for one variable need not be extreme for another variable." (Keilman et al. 2002)

The arguments above can be summarized in these points:

- Population modeling is usually deterministic, and could benefit from probabilistic modeling techniques.
- The deterministic model assumes that the underlying factors are perfectly understood.
- It is assumed that all of the contributing factors will occur as planned.
- Forecast models using multiple variables can compound the error due to assumed correlation.

3.3 County Borders

3.3.1 Naming Conventions

The county border shapefiles were obtained from the U.S. Bureau of the Census website and in the TIGER line file boundary format (U.S. Census Bureau 1997). The data set was complete and spatially accurate. The Federal Information Processing Standards (FIPS) code was used to join most of the other data as the primary key. The exception was the population table that did not include a FIPS key, instead it was necessary to use the county name concatenated to the two letter state abbreviation as the common field.

When the population table was joined with the shapefile there were a sizable number of null cells. Since the files came from different sources the County names did not follow the same naming convention. The only way to fix these errors was to manually edit the tables within ArcGIS. The most common problem occurred when the word "City" was omitted from a record in one of the data tables. For example, matching "Richmond" to "Richmond City".

3.3.2 Multi-polygons

The TIGER file standard creates multiple polygons for counties that have spatially disconnected regions (Figure 20). This issue is easily resolved, using the dissolve function of the Geoprocessing Wizard which combines the multi-polygon counties into single polygon.

Figure 20: An example of a single county with three distinct polygons (North Carolina)

3.4 Electricity Demand Data

The NEMS residential and commercial electricity demand forecasts have their own unique set of assumptions and input variables and the major points are discussed below (Energy Information Administration 2004b).

3.4.1 NEMS Residential Module Input Variables and Assumptions

The residential NEMS demand module projections are based on the number of households, size of households, number of appliances, and the energy efficiency of appliances as well as of the home. There is also an implicit assumption that there will be no radical changes in technology or consumer behavior although technology does change, both over time and in response to price change. Most of these assumptions are based on the Residential Energy Consumption Survey (Energy Information Administration 2003a). This survey of residential houses and their energy use acts as the jumping-off point for statistical analysis of the previous year's energy use data. Raw data such as household types per census division (Figure 21) are then aggregated into national scale data for analysis.

1999 RECS housing type survey by Census Division

Census Division	Single-family Units	Multiple family Units	Mobile Home	Total Units
New England	3.397,357	2,046,038	116,755	5,560,15
Mid Atlantic	9,022,447	5,618,800	376,390	15,017,637
East North Central	12,620,969	4,323,007	721,652	17,665,629
West North Central	5,729,603	1,659,511	389,346	7,778,460
South Atlantic	14,551,319	5,122,081	1,863,493	21,536,893
East South Central	4,751,956	1,205,518	795,918	6,753,392
West South Central	8,305,719	2,685,452	908,105	11,899,276
Mountain	4,912,205	1,601,455	560,142	7,073,802
Pacific	10,440,297	4,700,208	636,826	15,777,330
United States	73,731,872	28,962,070	6,368,627	109,062,569

Source: U.S. Deoartment of Energy, Energy Information Administration, *2001 Residential Energy Consumption Survey and Global Insight Macroeconomic Model T250803.*

Figure 21: The raw aggregate data used to create the residential demand outputs

Once the data were complied, two major adjustments were made. The first was the adjustment for the size of houses being built. House sizes are increasing at varying rates in different regions of the country. The second adjustment took weather and climate into account. The climate variable acted as the primary factor in the disaggregation model. Air conditioning and space heating are major contributors to electricity demand which is also intimately linked to geographic location. Climatic adjustments were made for the first three years of the forecast after which 30 year averages of Heating Degree Day and Cooling Degree Day were used (Energy Information Administration 2004b).

3.4.2 Commercial Module Input Variables & Assumptions

The NEMS commercial sector module relies on a very different set of variables and assumptions from the residential sector. First, it is important to detail what commercial energy entails as described by the EIA's Annual Energy Outlook (AEO),

> "the commercial sector includes business establishments that are not engaged in transportation or in manufacturing or other types of industrial activity (e.g., agriculture, mining or construction). The bulk of commercial sector energy is consumed within buildings; however, street lights, pumps, bridges, and public services are also included if the establishment operating them is considered commercial" (Energy Information Administration 2004b).

Since most commercial energy use occurs inside buildings, the module relies on data from the Commercial Building Energy Consumption Survey (Energy Information Administration 2002). Energy demand data for various types of commercial buildings (Figure 22) serve as the fundamental data upon which projections were built (Energy Information Administration 2004b).

1999 Total Floorspace by Census Division and Principal Building Activity
(Millions of Square Feet)

	Assem-bly	Educa-tion	Food Sales	Food Service	Health Care	Lodging	Large Office	Small Office	Merc/ Service	Ware-house	Other	Total
New England	378	575	10	40	86	169	565	311	824	429	348	3,735
Middle Atlantic	944	1,139	212	182	291	315	1,094	490	1,801	1,314	844	8,625
East North Central	1,202	1,506	115	463	336	725	1,096	847	2,183	1,983	751	11,205
West North Central	864	744	58	95	176	215	560	555	1,227	782	281	5,556
South Atlantic	848	997	156	302	312	825	1,507	1,077	2,611	1,909	457	11,001
East South Central	781	438	101	166	103	467	331	395	1,288	963	187	5,220
West South Central	1,028	913	135	207	215	303	663	644	1,569	1,085	501	7,264
Mountain	680	758	103	104	113	545	458	389	586	520	322	4,579
Pacific	1,074	1,580	105	292	233	956	1,145	969	1,698	1,493	607	10,152
United States	7,798	8,651	994	1,851	1,865	4,521	7,418	5,678	13,786	10,477	4,298	67,338

Note: Totals may not equal sum of components due to independent rounding.

Source: Energy Information Administration, Commercial Buildings Energy Consumption Survey 1999 Public Use Data

Figure 22: The raw aggregate data used to create the commercial demand outputs

The commercial sector module also makes assumptions about technological advances and their rate of adoption by businesses. These businesses can choose to retrofit old systems or build new ones at the end of their lifespan. By tracking changes in equipment, it is possible to project the types and efficiency of equipment being used each year.

There is also an adjustment for weather and climate in the commercial module, similar to that described in the residential module. However there will be different weighting schemes for each in order to account for the wide variations in the effects of weather and climate on the commercial and residential sectors.

3.4.3 Data Considerations

The NEMS output data for residential and commercial electricity demand was complied in separate .DBF (Access Database) files for input into the disaggregation model. The electricity figures of the NEMS outputs are measured in Quadrillion British Thermal Units (Quads). When the data are visualized (Figure 23), it becomes apparent that the distribution of the output data from these two modules are completely different. This fact is important to note for two reasons: first, there should *not be a comparison of the two modules* unless there is a comprehensive look at all end use categories to complete the picture, and secondly, there is a need for *two unique data classification schemes* to visualize the differentiation within the data set.

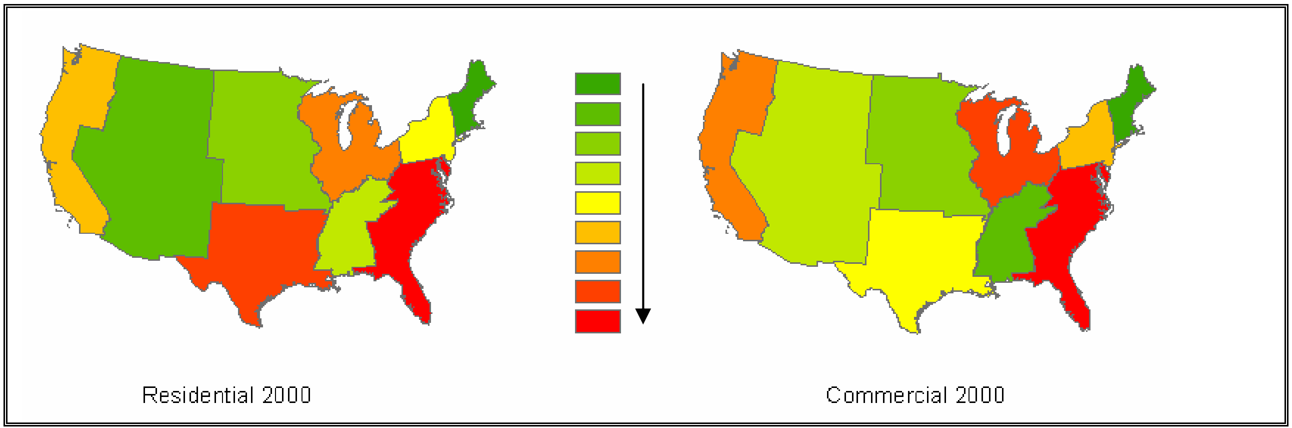


Figure 23: Total NEMS residential and commercial electricity demand per census division from lowest (green) to highest (red)

3.5 Climate Zone Data

3.5.1 Origin of Climate Zone model

Each of five distinct climate zones is defined by long-term weather conditions that affect the heating and cooling loads in buildings. The zones were determined according to the 45-year average (1931-1975) of the annual heating and cooling degree-days (base 65º F, 18.3º C). A region was assigned to a climate zone according to the 45-year average annual degree-days for its National Oceanic and Atmospheric Administration (NOAA) Division (Energy Information Administration 2004c). The climate zones were compiled using base data from the National Climatic Data Center (National Climatic Data Center 2004). Table 4 and Figure 24 and show the definitions of climate zones and conceptual method.

Climate Zone	Average Annual Cooling Degree-Days	Average Annual Heating Degree-Days
1	Fewer than 2,000	More than 7,000
2	Fewer than 2,000	5,500 to 7,000
3	Fewer than 2,000	4,000 to 5,499
4	Fewer than 2,000	Fewer than 4,000
5	2,000 or More	Fewer than 4,000

Table 4: The classification scheme used to derive the climate zones

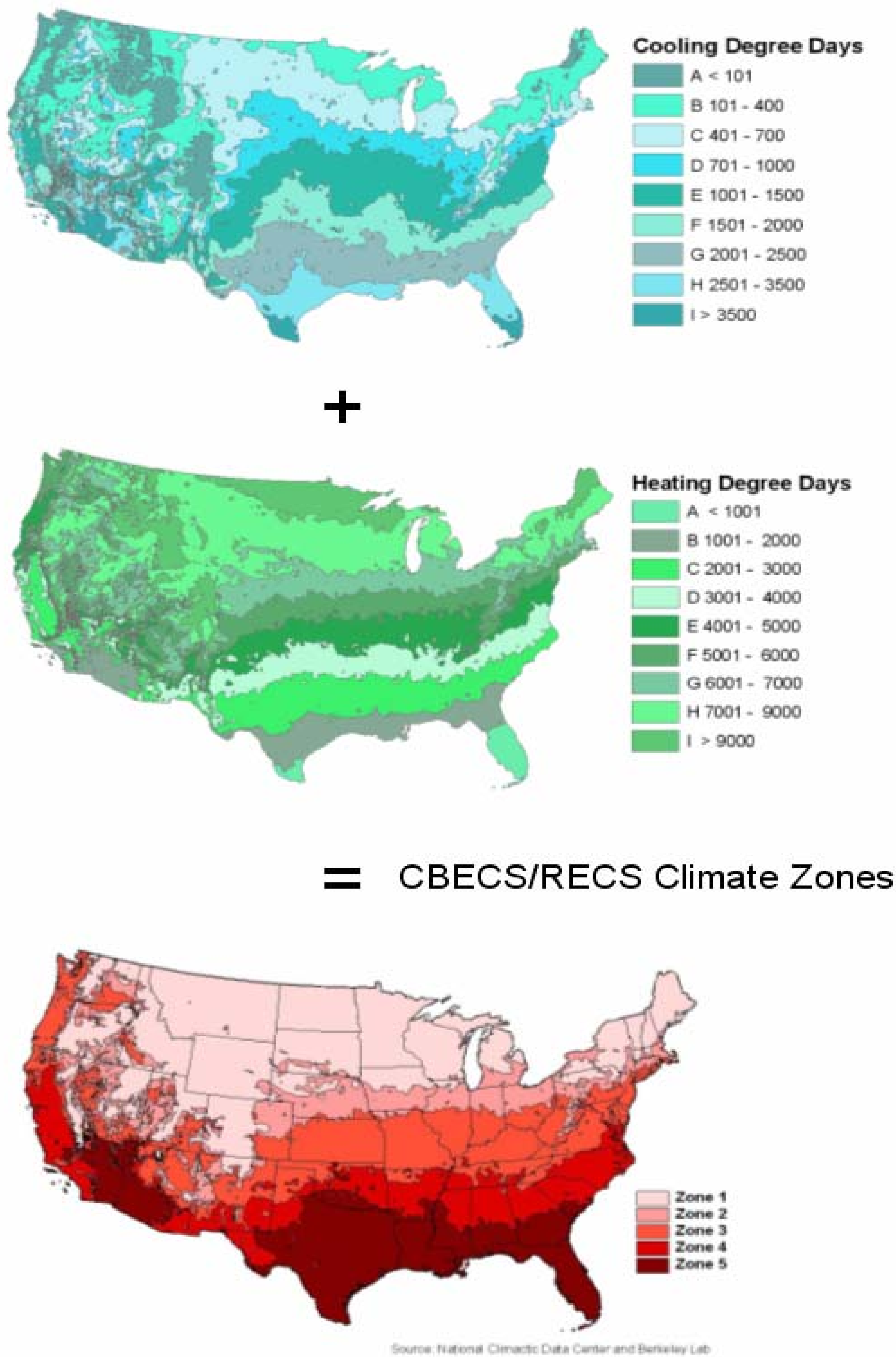


Figure 24: A conceptual model depicting how climate zones are derived using the HDD and CDD data

3.6 Comparative Testing Data

There is a scarcity of available information to test against the forecasted model. Limited state and county level data are available for actual electricity usage in the year 2000. Below are the sources and details for datasets against which the forecasts have been compared.

EIA provides official energy use data by state, broken down by end-use sector (Energy Information Administration). The data are compiled from retail sales of electricity to customers and are reported by the electric utilities and other energy service providers.

The California Energy Commission provides a collection of electricity use statistics by county in California (California Energy Commission 2002). These data are compiled from commercial sales of electricity to businesses and are reported by the electric utilities and other energy service providers.

4. Results

This section tests the validity of the disaggregation model using statistical and comparative procedures. First, comparison was made between the results of a population-based electricity demand model and the climate-weighted disaggregation model to determine the role of climate in making predictions on electricity demand. Second, a comparison was made between actual 2000 county by county electricity usage in California versus electricity usage predicted by the disaggregation model.

4.1 Rules of Thumb Regarding Forecast Data

Research into forecast accuracy yielded several conclusions (Garnet 1997) that form the basis for this chapter:

- Forecast accuracy generally increases with population size.
- Forecast accuracy generally increases for areas with slow, but steady positive growth rates, and decreases for areas with rapid population increases or losses.
- Forecast accuracy generally declines as the projection horizon (distance from the launch year) increases.

However, there are ways to compensate for these problems, such as:

Combine forecasts: Complete a number of projections and try to incorporate many of these in the final forecast. This assumes that every projection has error, but by completing and comparing different projections errors across these projections can be limited and a more accurate forecast is achieved.

Account for uncertainty: Use methods to incorporate the concept of uncertainty into forecasts.
Complete a range of projections using different assumptions (High, Med, and Low Series)
Use confidence intervals to generate different forecasts

4.2 Forecast Results of Disaggregation Model

Below are the disaggregation model outputs and comparisons, separated into state and county sections, and residential and commercial sectors[2].

Before looking at the disaggregation results below, it is important to discuss the nature of representing count data on variably-sized areas. For example, when comparing a large county with a high electricity demand, San Bernadino, CA, to a much smaller county with an equally high demand, Queens, NY, the absolute values may be similar but the demand per square mile (demand density) or demand per capita may be significantly different. Hence, the analysis of

[2] Alaska and Hawaii have not been added to these maps, but were a part of the analysis. They were omitted because of their distance from the continental US and their lack of influence on the output results. Larger versions of all result maps can be found in Appendix A: Map Results Gallery.

absolute demand values alone can be highly misleading. This section will underline the fact that there are various levels of analysis of electricity demand forecast and by combining these one can obtain a clearer and more meaningful picture of the underlying phenomenon.

The following methods were used to visualize the results. The electricity demand value and per capita maps are classified using natural bins or clusters. In this method each class will contain values that fall into natural groupings of data values and is useful in visualizing socioeconomic data.

The major assumption of the model is that climate will act as a key intervening variable affecting the demand for electricity, specifically in warmer regions. This assumption provides the basic premise for choosing and testing the model. Each section discusses the results and their relation to this basic premise.

4.2.1 Residential Forecast

The results for the time-series predictions of the residential disaggregation by county appear in Figure 25. The maps show a trend of high electricity consumption in the southwest region of the country and around the major cities. The entire east coast of the country is fairly saturated with only subtle increases over time. From a visual perspective one can see the higher electricity demand in warmer climates.

The results for the time-series predictions of the residential per capita disaggregation can be seen in Figure 26 which shows total demand per capita by county. This map shows a very different story from Figure 25. There is a direct correlation between the census divisions and the distribution. Most important to note only in the southeast region do significant changes in per capita usage happen over time. The rest of the country does not vary significantly.

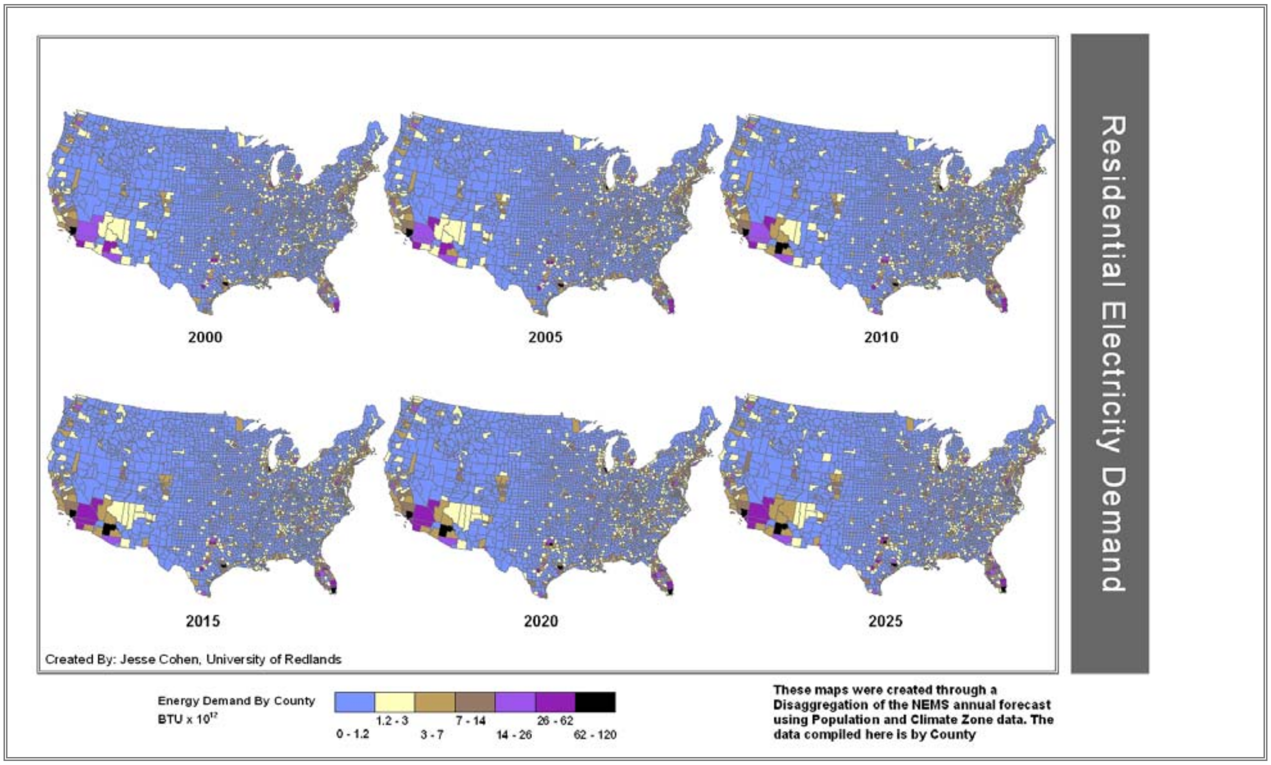


Figure 25: Residential electricity demand for the time period 2000-2025

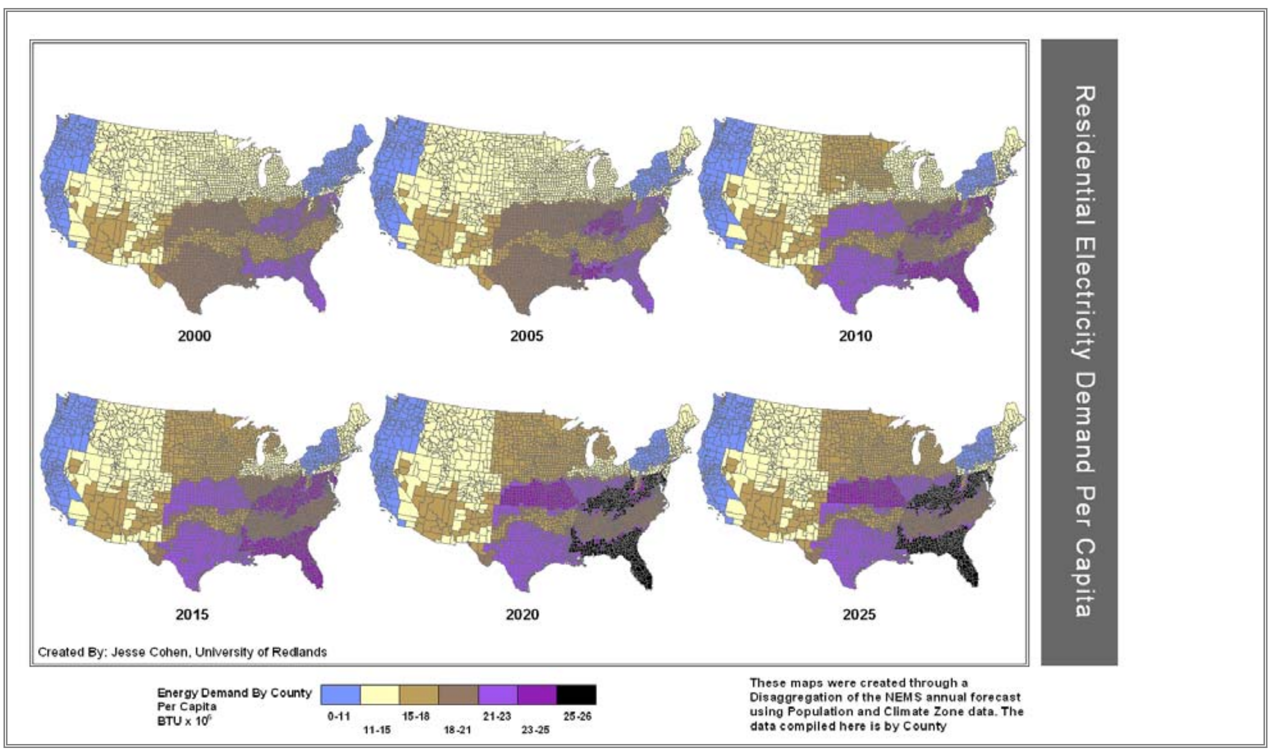


Figure 26: Residential electricity demand per capita 2000-2025

4.2.2 Commercial Forecast

The results for the time series predictions of the commercial disaggregation model can be seen in Figure 27. The results show a predictable increase in electricity demand radiating from the major cities. The central region of the country has little variation with the majority of the counties falling in the lowest category of electricity demand. Areas of change to note are the southwest and Texas region which are predicted to experience intense growth of both population and electricity demand over the 25 year period.

The results for the time-series predictions of the commercial disaggregation model can be seen in Figure 28 which shows total demand per capita by county. This map provides a different perspective from Figure 27. Between 2000 and 2025 change in per capita demand begins in the southeast and continues north and west. By 2015 the entire south has increased its per capita demand significantly. It is not until 2020 that the northern regions values begin to increase and the southwest becomes a major source of demand. The only region of minimal increase over the 25 year period is the west.

A comparison of the Pacific region in particular provides a clear example of the value of using per capita data on a county basis. When looking at the per capita values of the Pacific region it appears that there is more efficient electricity usage than in other parts of the country, such as Florida. In contrast, when using the county value analysis the electricity usage appears highly concentrated in the Pacific region. This vital difference could not be determined unless both of these analyses were conducted.

From the viewpoint of public policy both of these approaches provide valuable information regarding electricity usage. County by county data provided absolute values for electricity demand which can be used for planning changes in regional electricity infrastructure, whereas per capita data provide information that can be used to assess the regions in which electricity efficiency standards can be improved. Hence, both forms of analysis are required to obtain a complete picture of the regional electricity usage.

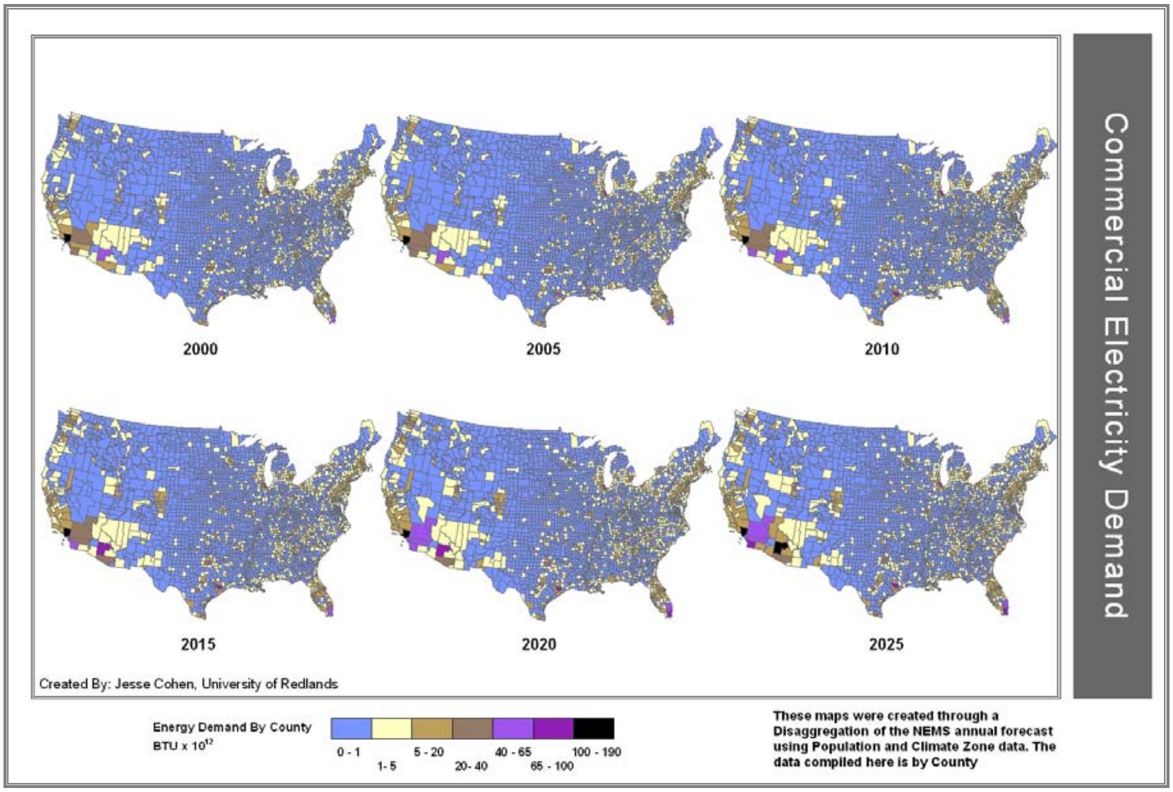


Figure 27: Commercial electricity demand for the time period 2000-2025

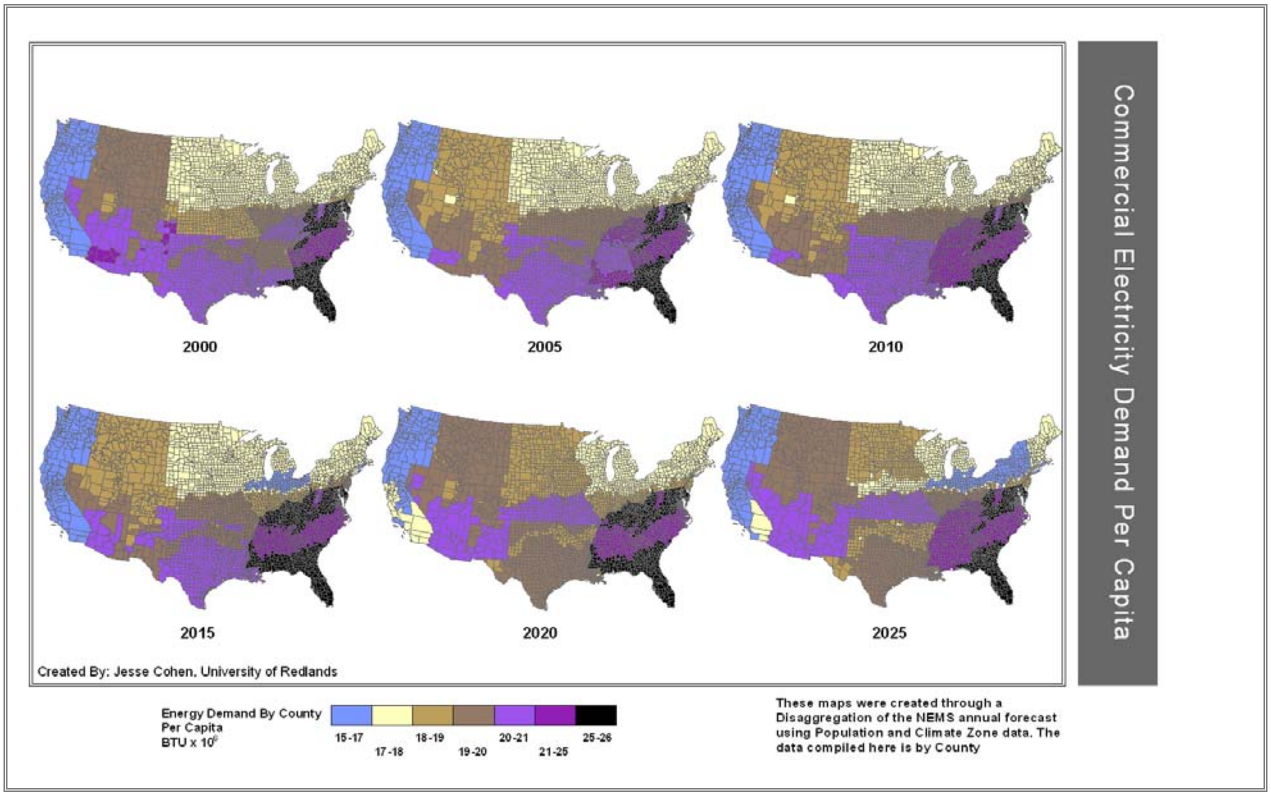


Figure 28: Commercial electricity demand per capita by county for 2000-2025

4.2.3 Residential Module: Percent Change Over Time

The residential electricity growth rate indicates most counties growing at the rate of 1-40% between 2000 and 2025 (Figure 29). The middle of the country was the only area that showed a negative rate of growth and those counties were centered in North Dakota, Nebraska, Louisiana, and Missouri. Both Florida and Arizona showed a number of counties with growth in the 95+ % range. These numbers corresponded to the trend of people moving to the southwest in particular, and southwards in general[3].

---

[3] Both South Dakota and Iowa appear on the map completely colored red. This artifact is a result of the forecasting method used on the states with no forecast data (see Section 3.2).

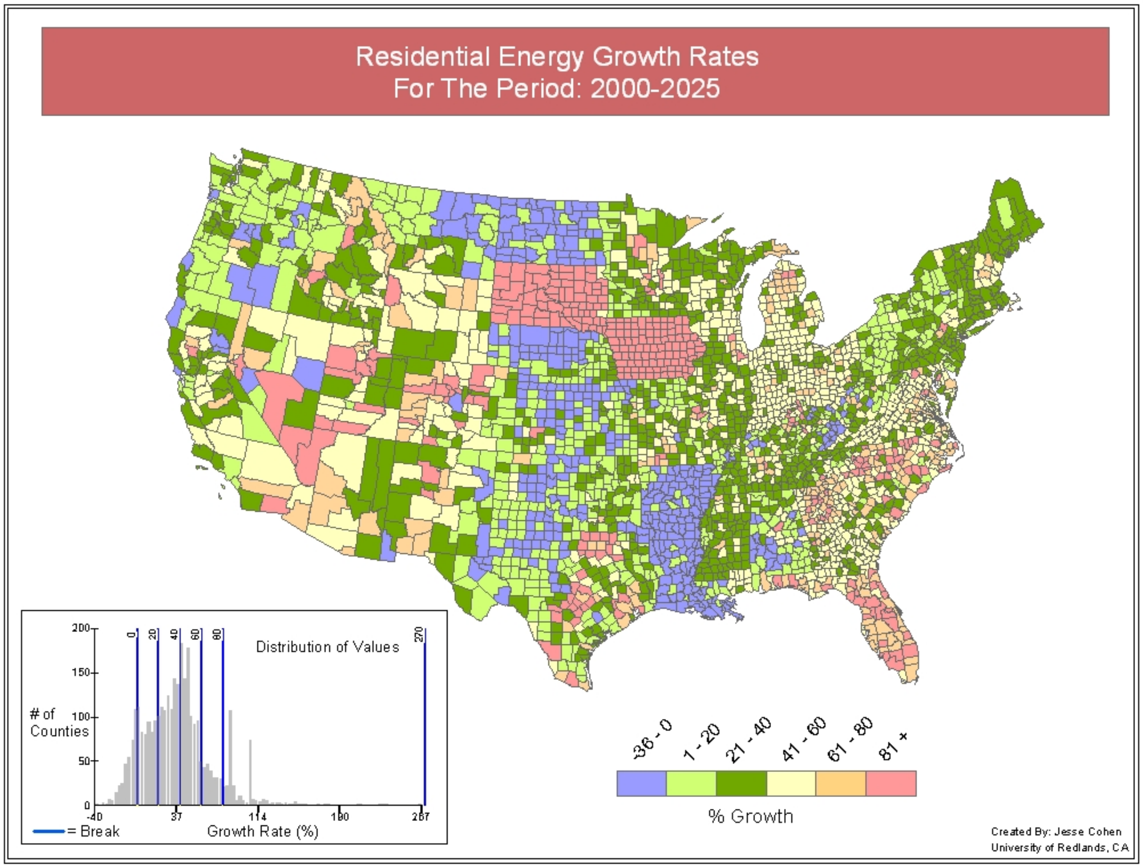


Figure 29: Residential electricity growth rate (percentage) from 2000 to 2025

4.2.4 Commercial Module: Percent Change Over Time

The commercial module exhibits some important differences when compared to the residential module. The distribution of energy growth was similar but the numbers were elevated. The entire west and south Atlantic regions showed large numbers of counties at the highest growth rate (81+ %) over the 25 year period. The middle of the country was again the only region with a negative growth rate. Texas showed a high growth rate in its eastern half and simultaneously a low growth rate in its western half (Figure 30). The results clearly showed a significantly higher rate of change in the commercial sector compared to the residential sector.

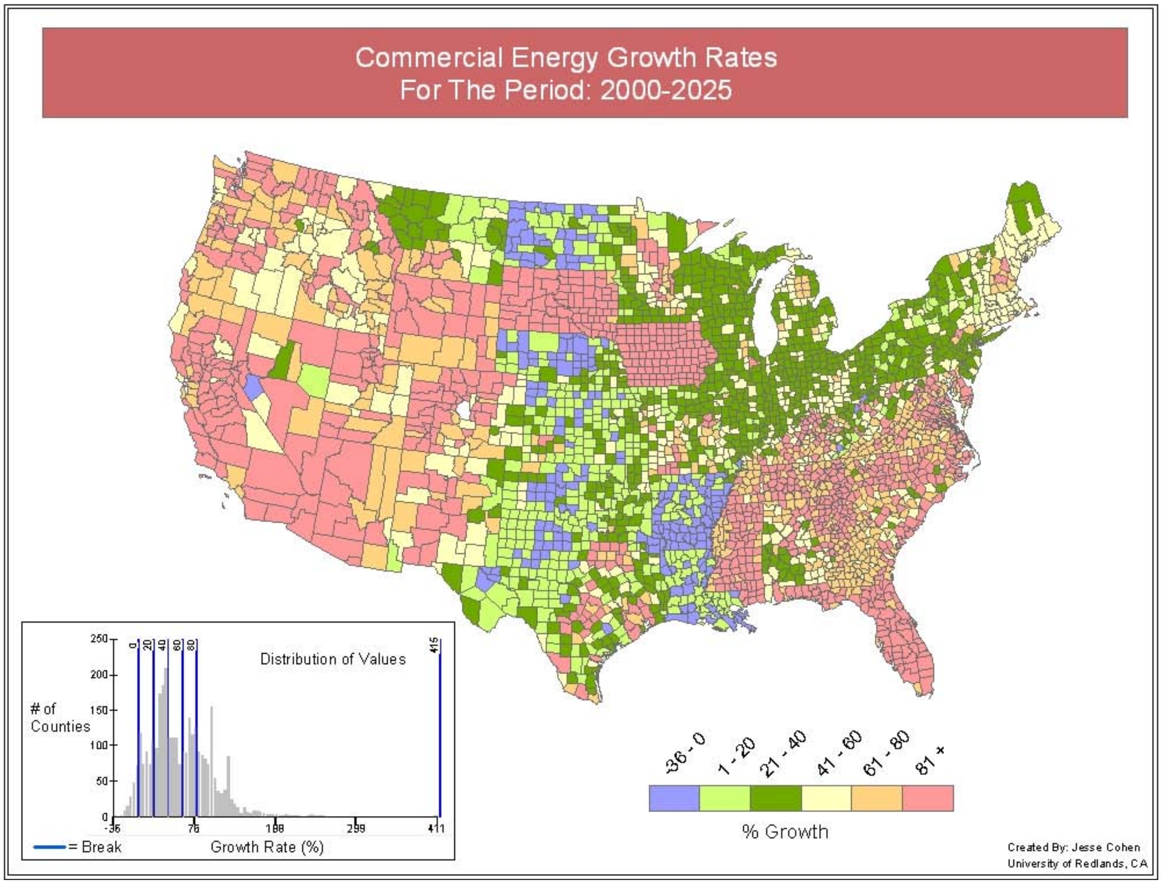


Figure 30: Commercial electricity growth rate (percentage) from 2000 to 2025

4.3 State Level

The disaggregation output data were compared to State Energy Consumption 2000 values (Energy Information Administration 2004d). To avoid the varying county size issue (Section 4.2) both data sets were normalized by population to make per capita maps. The electricity usage figures from DOE were compiled by state for the major end-use consumers. These values were classified and mapped for comparison against the predicted outputs, aggregated to the state level. The results were classified into high, medium and low categories, a result of breaking the data into three classes each with an equal number of features.

The predicted electricity demand values were also used to calculate the percent variation from the actual electricity demand data. The percent variation map in the lower left-hand corner of the state per capita demand maps indicate how inconsistent the predicted values were. The main point to be emphasized here is that the most complete picture of electricity demand can best be described by using both the electricity demand per capita and percent variation maps in combination.

4.3.1 Predicted Forecast Variation: Residential

Upon analysis, considerable variation was found between predicted demand and actual electricity usage. Of the 48 states analyzed, it was found that the forecast was inaccurate for 15 of the states to the extent that the predicted values were +/- 30% of the actual values (Figure 31).There was very little geographic pattern to the under-predicted states locations. The over-predicted states were grouped in the middle of the country in relatively unpopulated areas. In the case of the Pacific region, Washington and Oregon had a disproportionately high actual electricity use per capita, possibly due to extensive use of electric heating.

In conclusion, the prediction model was successful in the moderately-to-highly populated, temperate regions of the country. The model was less accurate in the warmer and less-populated areas, creating high levels of variation from the actual map.

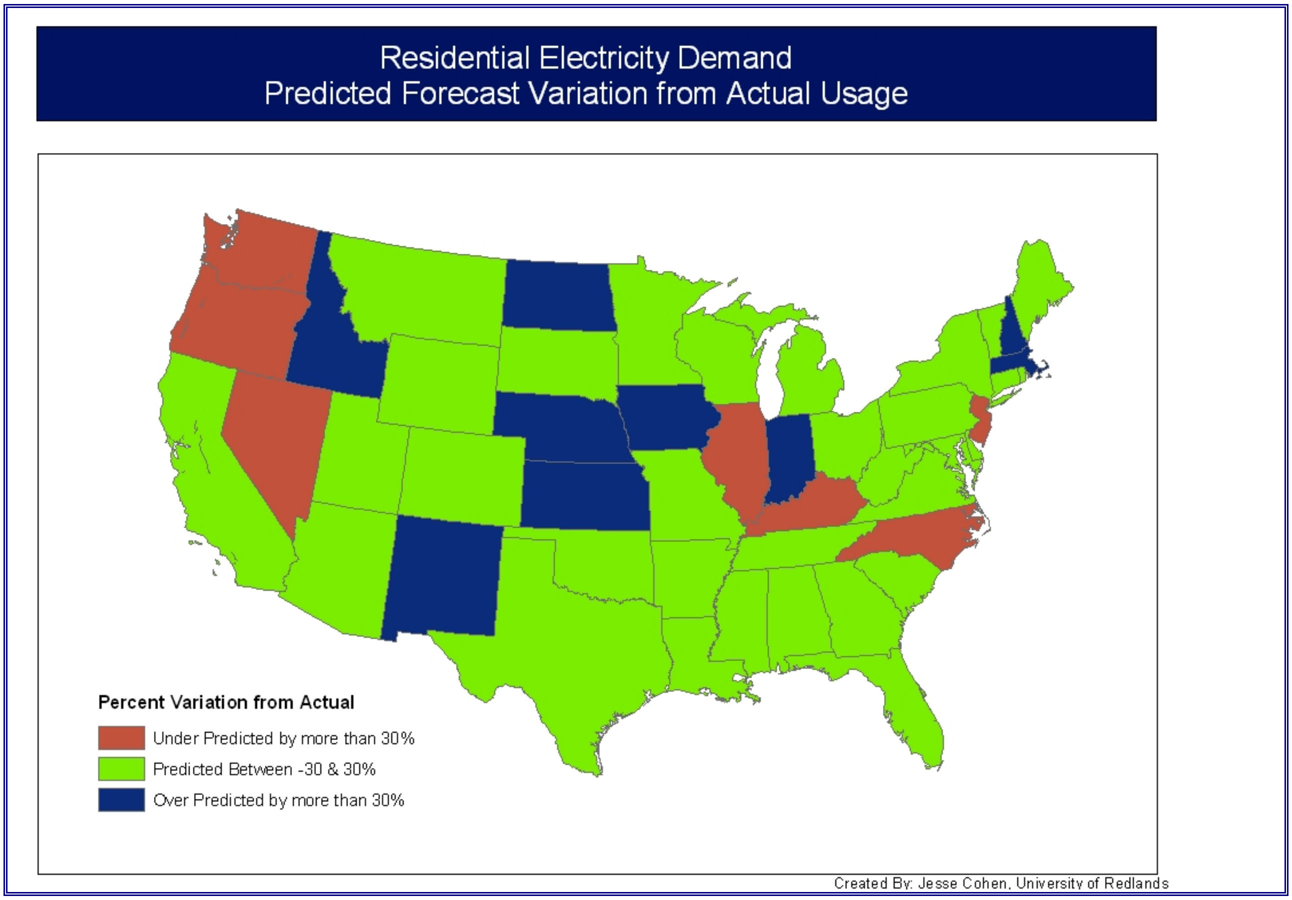


Figure 31: The percent variation of the predicted demand values from actual 2000 data.

4.3.2 Predicted forecast variation: Commercial

Similarly, upon analysis, considerable variation was found between predicted demand and actual 2000 electricity usage. Of the 48 states analyzed, it was found that the forecast was within +/- 30% of actual values for 24 states (Figure 32). There was no discernable geographic pattern regarding the location of the under-predicted states. However, the majority of over-predicted states were grouped in the middle of the country in relatively unpopulated areas. In conclusion, the prediction model accurately predicted year 2000 results in the moderately-to-highly populated regions but began to break down in the less populated states creating high levels of variation from the actual data.

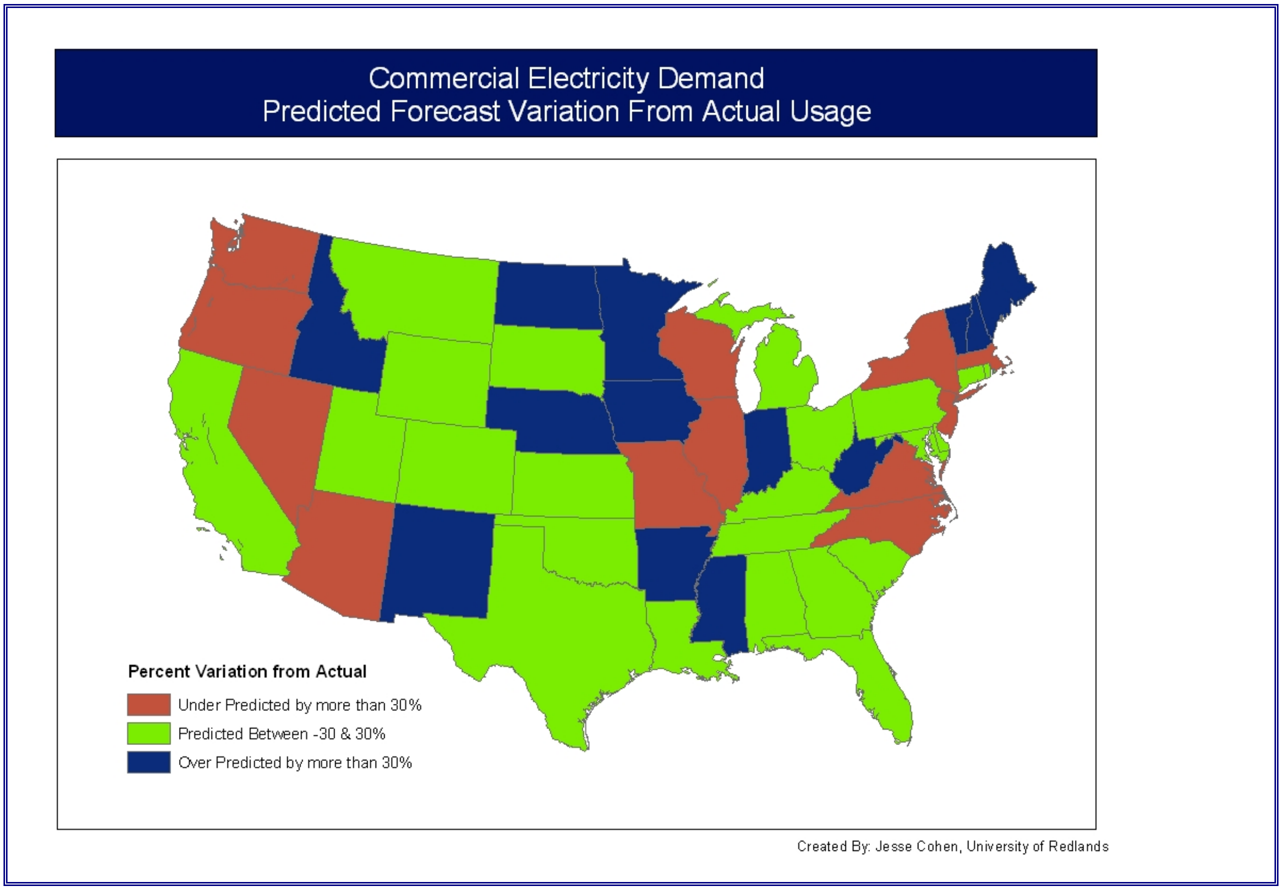


Figure 32: The percent variation of the predicted demand values from actual 2000 data.

4.4 County Level

California is one of the few states that compiles electricity use by county and makes it available to the public. Unfortunately, the data are only available using Residential/non-Residential classes making a comparative test only possible for Residential electricity use.

4.4.1 Predicted Forecast Variation: Residential Only

A principle problem with this data was the fact that the electricity demand forecast for the entire State of California differed notably from the actual demand (Actual 277 x 10^{12} Btu vs. Predicted 344 x 10^{12} Btu). County level forecasts tend to be the least accurate because of their high level of resolution compared to regional forecasts (Garnet 1997). The map of the regional distribution of percent differences showed a pattern of over prediction in the southern counties and under prediction in the northern inland counties (Figure 33).

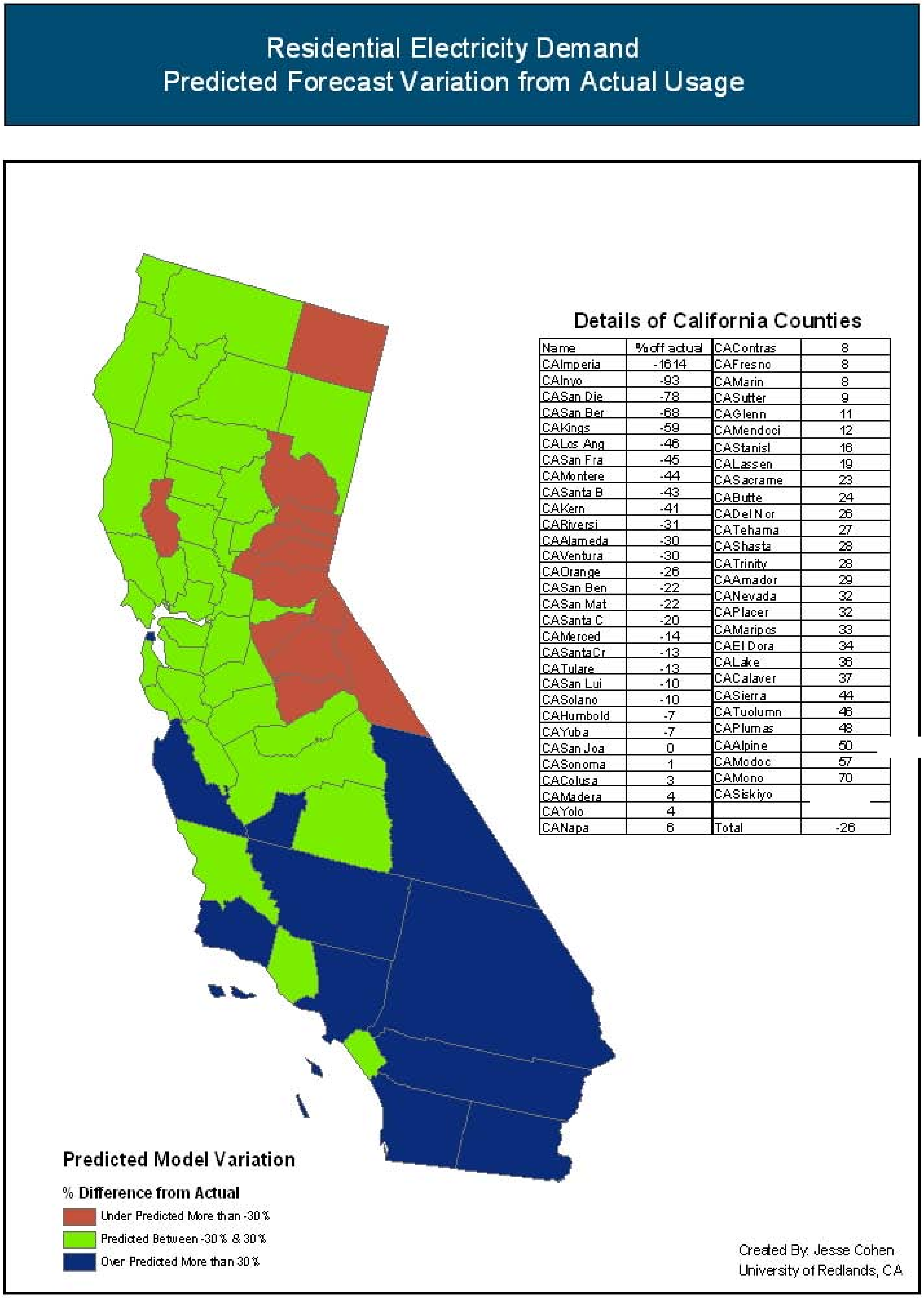


Figure 33: The 2000 California counties predicted model's variation from the actual model in terms of percentage over or under estimated.

The working assumption that warmer more populated counties would consume higher levels of electricity per capita was refuted by these data.

The over estimation could be a product of the climate zone scaling factors that were derived from the RECS. It also could be a lack of precision of the delineations of the climate zones.

In conclusion, the prediction model was successful in the moderately populated, temperate regions of the state. The model lost its predictive value in areas with warmer climates with high and low populations (see the percent variation tables in Appendix B).

4.5 The T-Stat Test

Since the output from the prediction model is highly correlated with the input population data it is important to compare the two to examine whether their influence on the disaggregation outputs was significantly different. The T-stat test compares two sets of values to determine whether or not they are significantly different. To create a population-based disaggregation for comparison, Excel was used to calculate a proportion between the population for each region and the electricity demand. First, the county populations for each region were divided by the population sum of the region and multiplied by 100. Then the percentage of the region population for each county was multiplied by the region's electricity demand forecast. This assigned each county a value for electricity demand that was proportional to their population (Table 5). These outputs were useful for testing the predictive model, which incorporates the climate zones, and seeing if the results were statistically different.

Region 1 --- Total Energy Demand: 600 units

	Population	% of Region	Energy Demand
County 1	75 / 300	25%	25% * 600 units = 150 units
County 2	75 / 300	25%	25% * 600 units = 150 units
County 3	150 / 300	50%	50% * 600 units = 300 units
Total	300	100%	600 Units

Table 5: Population based model explanation

Two samples for means test allows two dependent variables to be compared for significant differences. A sample of 30 random counties was analyzed in five separate versions of the test and the conclusion in all the tests was that the two methods were *not* significantly different. Since the main variation in temperature should have a relationship to the location of each state the T-stat test was then performed to test each state's individual variation by aggregating values to the state level.

The map in Figure 34 shows the results of the T-stat tests for each state. The states in blue represent those in which the outputs of the two models were significantly different, while the red states represent the outputs that were not.

These results <u>did not conform to the original hypothesis, ie., that climate would be a major variable in determining electricity demand</u>. The assumption was that the electricity demand in southern, warmer states would be affected more by climate in relation to the population. The data showed that in the southern regions, population alone is as good an indicator of electricity use as the model in which climate zone data is factored in with population data. The northern, colder climates appear to be more accurately described using the two factors. The main point that can be distilled from the T-stat analysis is that there are different variables influencing electricity demand among various regions of the country.

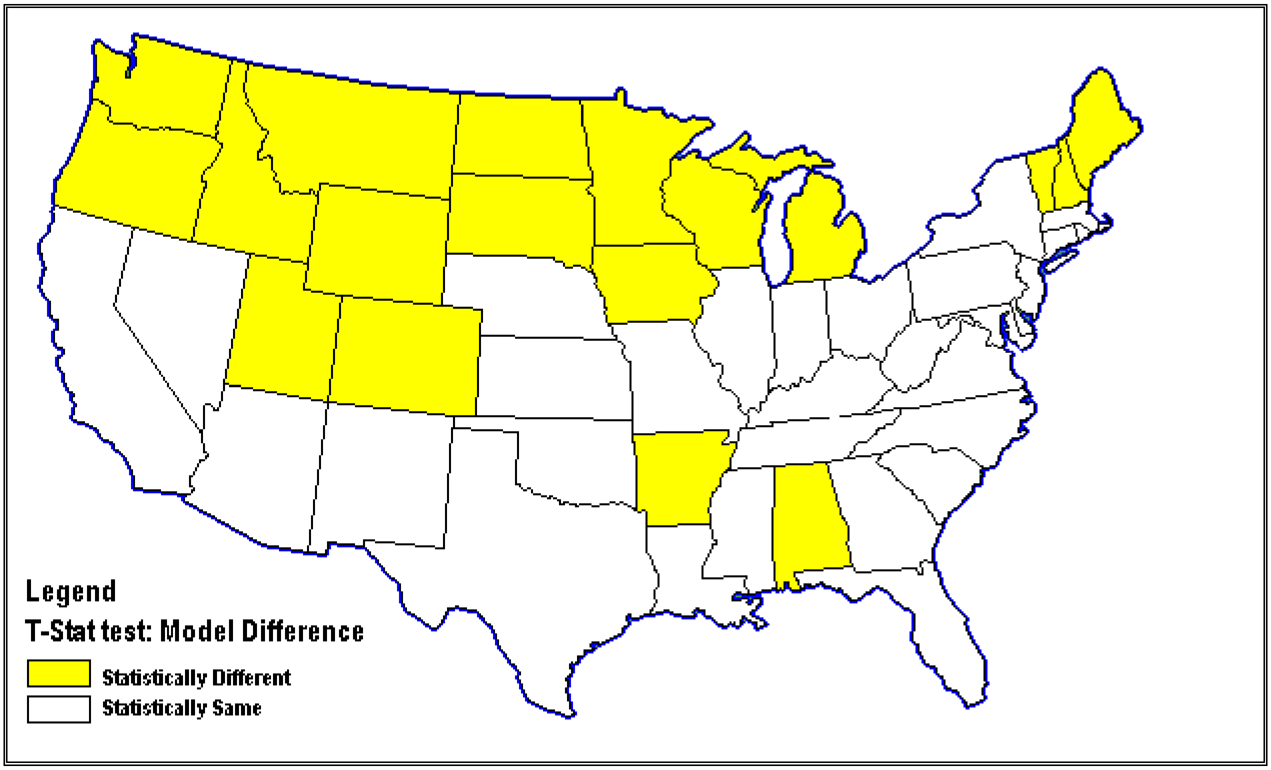


Figure 34: Statistical differences between disaggregation outputs and population

4.6 Summary

The forecast results of the disaggregation model can be summarized as follows:

Residential Electricity Demand

- Both the predicted and actual data showed electricity use to be highest in southern states and lowest in the northeast.
- The greatest increase in electricity demand per capita over the next 25 years is predicted to be in the southeast.
- The highest growth rates of demand were found in the south atlantic, east north central and southern pacific and mountain regions.
- Fifteen states' predictions for the year 2000 varied more than 30% from the actual electricity demand data.

- The T-stat test showed no overall significant difference between the results from the disaggregation and population-based forecasts.
- The state by state T-stat tests displayed a trend of northern states having significant differences in model results, while the southern states showed no significant difference.

Commercial Electricity Demand

- Electricity demand is highest in southern states
- The greatest increase in electricity demand per capita over the next 25 years will be in the southeast.
- The highest growth rates of demand were found in the south atlantic, east north central and southwest and mountain regions.
- Twenty-four states' predictions fot 2000 varied more than 30% from the actual electricity demand data.
- Some states had a disproportionately high electricity use in comparison to their population (e.g. Washington, Oregon)

The results of the forecast model were counter intuitive in that in the northern states, climate data was more predictive than in the southern states. This suggests that there is a flaw in the assumptions made in creating the disaggregation process. However, although air conditioning usage is driven by climate, it is also true that other end uses of electricity are climate driven. For example, people are indoors more in colder climate driving up lighting and other end uses. Possible model improvements could include: (1) either more or better influencing factors added to the disaggregation of electricity demand data, or (2) a better set of weighting factors by which to derive the influence of climate in the various census regions.

5. Time-Series Animation

The results of the disaggregation process are useful in many forms. The temporal aspect of the outputs cannot be fully visualized in a static medium. Using Macromedia Flash, the time-series can come alive as a visual representation of the change of electricity demand. The animated movie medium is familiar and makes it possible for a wide audience to easily absorb this complex and rich data.

5.1 Data and Cartography

Residential and commercial electricity demand per capita forecasts were the primary data input into the flash environment. The data were split into three types for varying perspectives of visualization:1. total values per county, 2. total density per square kilometer, and 3. per capita values by county. The reason for this was that the density and per capita calculations represented change in a significantly clearer and more accurate manner. When the raw data was mapped it was harder for the eye to determine the rate of change. This had to do with the mapping count data over a variable area as described earlier in Section 3.2. To overcome this issue, the original value maps were replaced with density (sq km) and per capita maps.

5.1.1 Classification

The classification scheme originated from the Natural Breaks method. Both the residential and commercial data required their own scheme due to the vastly different values and distribution. The basic scheme included 7-8 classes to cover all of the main county sizes, and a rounding of the values to create an easy to read legend. The manually classified values were also shifted slightly lower to bring out the distribution among the smaller to mid-sized counties which represent the bulk of the counties in the USA.

5.1.2 Colors

The color scheme used for the raw data was a custom discreet color progression. The reason for this scheme was to show the vast majority of the counties in a lighter color, but to also show each of the other classes in a color that could be seen according to the number and size of the corresponding counties. The highest value range had only a few counties and required the most prominent color, and each color had to be clearly different from the majority of its neighbors.

5.2 Results

5.2.1 Electricity Demand Density Maps

The Density map (see Figure 35) gives the user a good sense of the changes occurring over time. For reference purposes it offers a button to see the 50 largest cities in the country on top of the data for visual associations. The state boundaries are overlayed for the sake of familiarity.

These density maps were created using the spatial analyst extension's density function. This function requires that the county centroids be attributed with the electricity demand values. In a simple density calculation for each grid cell, the centroids that fall within the defined search area are summed and then divided by the search area size to get each cell's density value. The search radius used here of 300 miles was a compromise between the high density Northeast and low density Pacific regions.

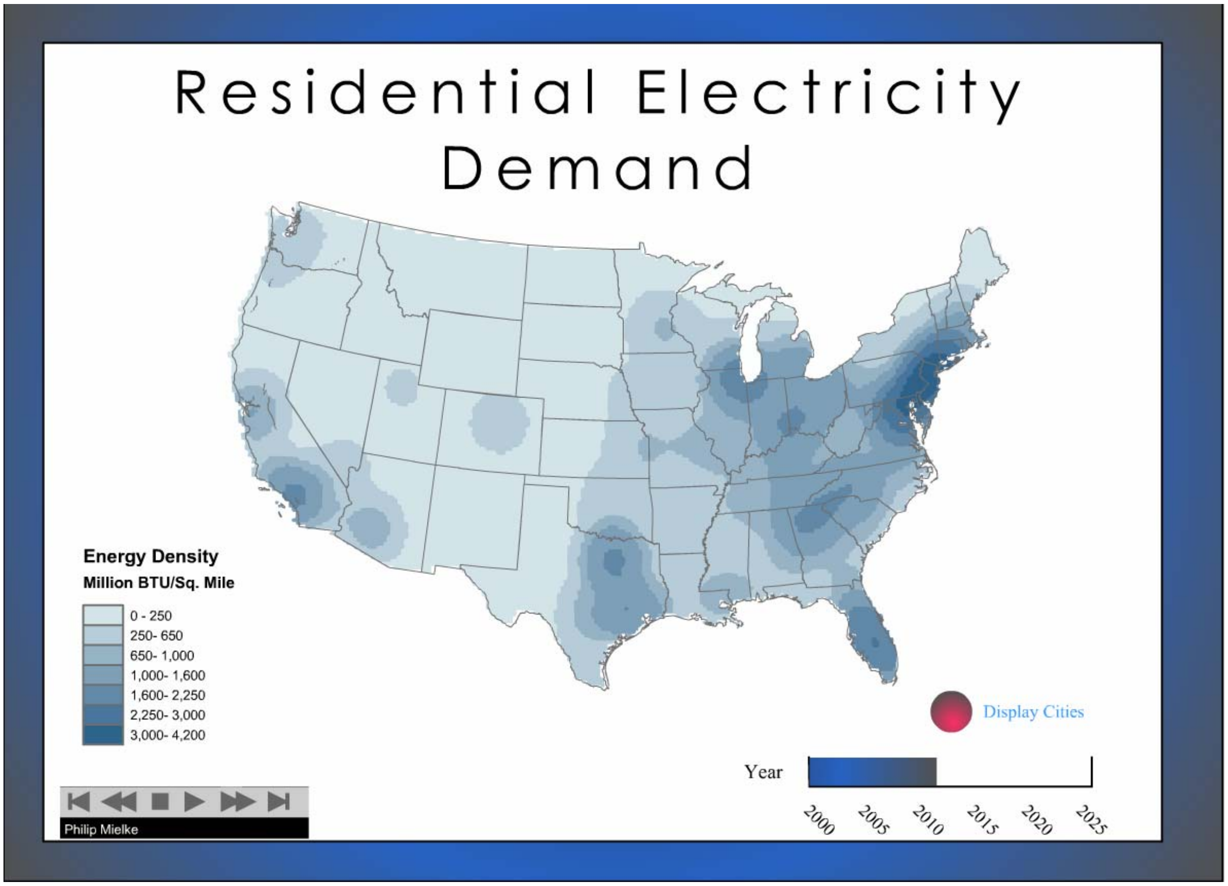


Figure 35: Macromedia Flash time-series snapshot of residential electricity demand density

5.2.2 Electricity Demand Per Capita Maps

The electricity demand by county forecast map was problematic to visualize. The U.S. counties range in size from San Bernardino, CA (32,468 km^2) to Manassas Park City, VA (2.7 km^2). Representing these variations in data required a deviation from some of the cartographic conventions. The main consideration was the county lines, and their width in comparison to their overall area. The county lines were drawn as thin as possible and in a gray color. The idea was to take the emphasis off of the county boundaries and onto the data within the counties. The map fulfilled this objective in conjunction with additional information of "Major Cities". The Major Cities button displays the 50 largest cities overlayed on the time-series map and can be switched on and off as a visual association aid (Figure 36).

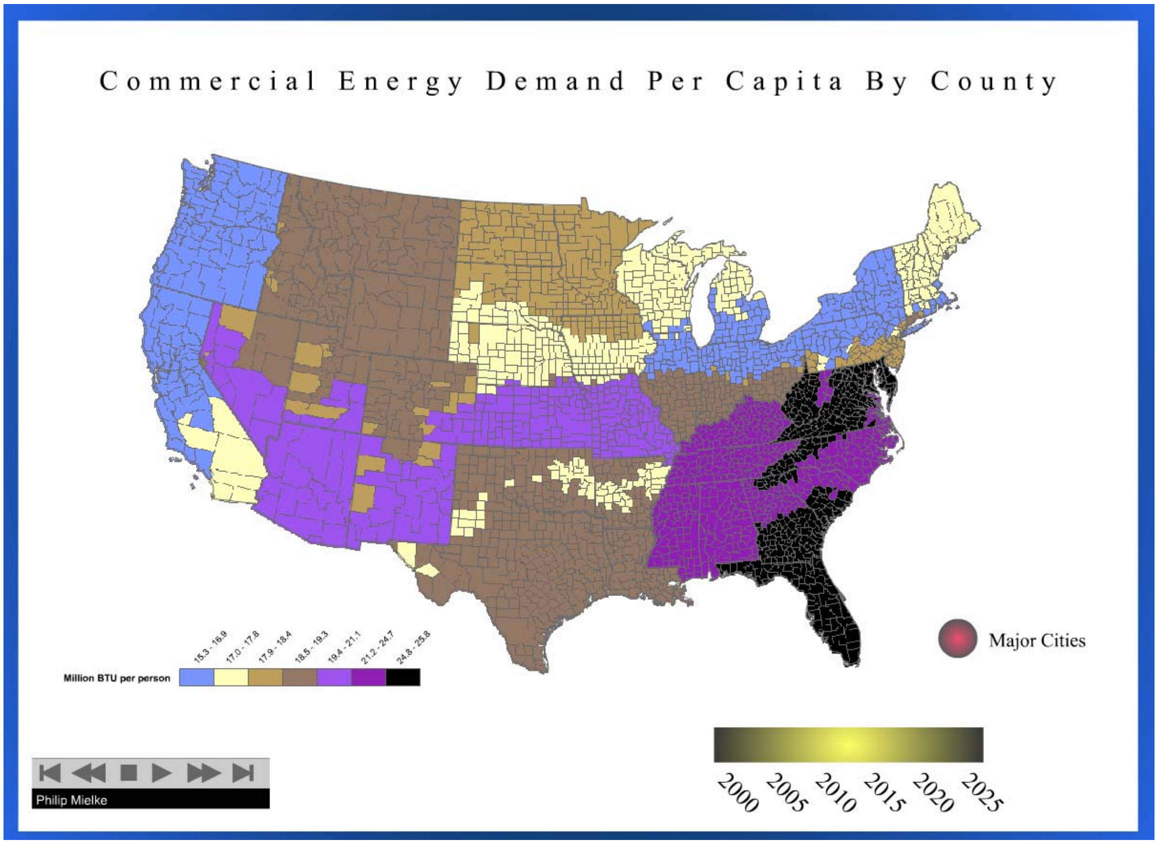


Figure 36: Macromedia Flash time-series snapshot of commercial electricity demand per capita

6. Conclusion

Spatial disaggregation can prove to be a valuable tool when building policies and programs based on national data compilations. There are a great number of datasets that could benefit from more detailed levels of resolution. As methods become refined and statistical techniques are integrated with GIS, a solid set of fundamentals will emerge.

There were four main goals of this work:

1. To compile population forecast data to the year 2025
2. To identify a disaggregation method to be used on NEMS data
3. To automate the disaggregation process within a user-friendly framework
4. To generate animations of electricity demand change from 2000-2025.

The population data input into the disaggregation model was compiled with the aid of statistical techniques. The population data were useful for the purpose of testing the disaggregation method, but have severe limitations. Each county's projection was calculated independently using different methods and assumptions. Another problem is the statistical inferences needed to complete the incomplete or null data records.

The methods which were developed to disaggregate the NEMS output followed established techniques of Areal Interpolation and utilized existing public domain data. The choice of method Areal Weighting using Control Zones was driven primarily by the nature of the problem. Ideally, alternative methods could be tested and assessed before a final model is selected for automation.

The NEMS Disaggregation Engine tool was successfully created meeting all of the criteria outlined in the user-needs analysis. The tool gives users the ability to disaggregate NEMS data through the use of a user-friendly Graphic User Interface (GUI). This GUI should make the usage of the NEMS data more accessible to analysts and other professionals with common interests. The tool also gives users flexibility to experiment and create different scenarios quickly and easily. For example, the tabular data can be transformed into Flash animations for ease of display via the web and can be an inexpensive alternative to GIS for representation of change over time.

The goal of this work was to demonstrate the utility of applying GIS techniques, including spatial analysis, to the field of electricity analysis. Although further enhancements are still necessary, making the transistion toward a spatial perspective is a significant step in regional electricity analyses conducted on a country-wide scale to better represent this spatial data.

7. Future Work and Direction

Future possibilities for this work include: 1. population forecasting, 2. disaggregation methodology, 3. electricity analysis/GIS integration, and 4. animation of time-series data for visualization purposes. Each of these has potential for further development and are discussed in detail in this section.

7.1 Population Forecast

The first and most straight-forward improvement would be the addition of a more accurate county population forecast. A consistent and accurate forecast would allow analysts the chance to determine the ranges of error and incorporate those errors into the model design. Currently, it is not possible to determine the error since the forecasting methods are unknown.

7.2 Disaggregation Methodology

Alternative methods for the disaggregation of NEMS data could be tested. There is strong support for the use of raster grids for this type of analysis. This approach could take other factors into account in addition to climate zones to compile an accurate set of variables influencing electricity demand. This would also allow the inclusion of a masking grid to eliminate regions of uninhabited lands and water bodies. This masking would significantly change the density of electricity demand in counties with large land areas but small regions of habitable land.

7.3 Electricity Analysis/GIS Integration

One of the main purposes of creating the electricity demand forecasts is to allow analysis of the data in relation to electricity supply models. This requires integrating PowerFlow models, used by the electricity analysis community, to create networks and compile important data at points in the network, i.e., generation sites, substations, and high-voltage power lines. The integration of these two platforms has yet to be standardized and is still in the experimental phase. With more time and resources, including access to the PowerFlow modeling software, there could be further study into the outputs of the NEMS disaggregation in comparison to the outputs of the PowerFlow models to determine future regions of high demand/low supply. These regions would ultimately be overlayed with alternative electricity data to determine regions where distributed generation would be a worthwhile investment.

7.4 Animation of Time-Series Data

The representation of time has always been an issue in the field of GIS. Macromedia Flash is capable of animating maps to show changes over time. ESRI has also released an extension called Tracking Analyst that offers some level of time animation though it was not applicable to this project.

The advantages of Flash are the ease of use and the ability to create a unique interface with buttons and windows. Another advantage is the ease of transference to the World Wide Web. Flash can be inserted into an ASP/JSP page for a GIS/IMS implementation and run as a frame or as the entire website. This functionality offer a wealth of opportunities including comparison studies of different modeling techniques, as well as the addition of dynamic multimedia links to data.

References

California Energy Commission. 2002. *2002 - 2012 Electricity Outlook Report*. P700-01-004F. February. Sacramento, CA.

Deichmann, U. 1996. *NCGIA Technical Report: A Review of Spatial Population Database Design and Modeling*. UC Santa Barbara. Santa Barbara. 1-62.

Energy Information Administration. *Annual Energy Review 2004*. DOE/EIA-0384 (2004). Washington, DC.

Energy Information Administration. 2002. *1999 Commercial Buildings Energy Consumption Survey - Detailed Tables*. EIA/DOE. August. Washington D.C.

Energy Information Administration. 2003a. *2001 Residential Energy Consumption Survey - Detailed Tables*. EIA/DOE. October. Washington D.C.

Energy Information Administration. 2003b. *The National Energy Modeling System: An Overview 2003*. EIA/DOE. March 2003. Washington D.C. 72.

Energy Information Administration. 2004a. *Annual Energy Outlook 2004*. U.S. Department of Energy. DOE/EIA-0383(2004). January 2004. Washington D.C.

Energy Information Administration. 2004b. *The Assumptions of the Annual Energy Outlook 2004*. DOE/EIA-0554(2004). February. Washington, DC.

Energy Information Administration. 2004c. "Energy Glossary Website." *http://www.eia.doe.gov/glossary/*.

Energy Information Administration. 2004d. "State Energy Data 2001: Consumption Estimates." *http://www.eia.doe.gov/emeu/states/sep_use/total/csv/use_csv.html*.

Garnet, S. 1997. "Population Forecasting: What Method is Best?" Florida State University. *http://garnet.acns.fsu.edu/~tchapin/urp5261/lectures/Choosing%20a%20Forecasting%20Method.ppt*.

Goodchild, M., L. Anselin, and U. Deichmann. 1993. "A Framework for Areal Interpolation of Socioeconomic Data." *Environment and Planning A*, 25, 383-397.

Keilman, N., D. Q. Pham, and A. Hetland. 2002. "Why Population Forecasts Should Be Probabilistic - Illustrated by the Case of Norway." *Demographic Research*, 6 (15), 410-449.

Leddy, R. 1994. "Small Area Populations for the United States." *Assocation of American Geographers Annual Meeting*, San Francisco.

National Climatic Data Center. 2004. "NCDC's Data Access Tools." National Oceanic and Atmospheric Administration. *http://lwf.ncdc.noaa.gov/oa/dataaccesstools.html*.

Norman, P., and C. Marsh. 2003. "Dealing with Changing Boundaries." *Use of GIS to Understand Socio-Demographic and Economic Change: A Workshop*, Manchester, England.

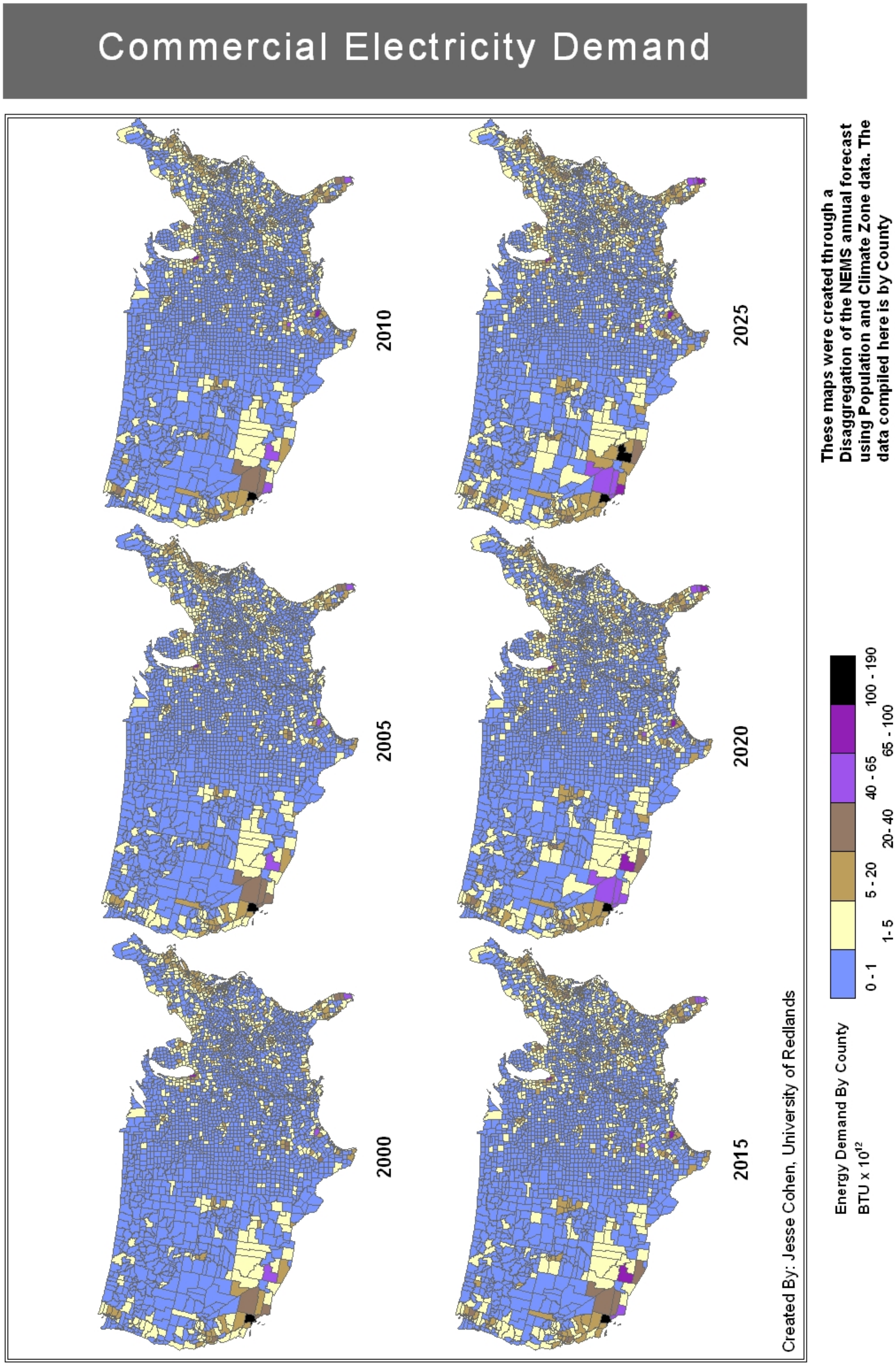
U.S. Census Bureau. 1997. "Topologically Integrated Geographic Encoding and Referencing (TIGER) Products." *http://www.census.gov/geo/www/tiger/*.

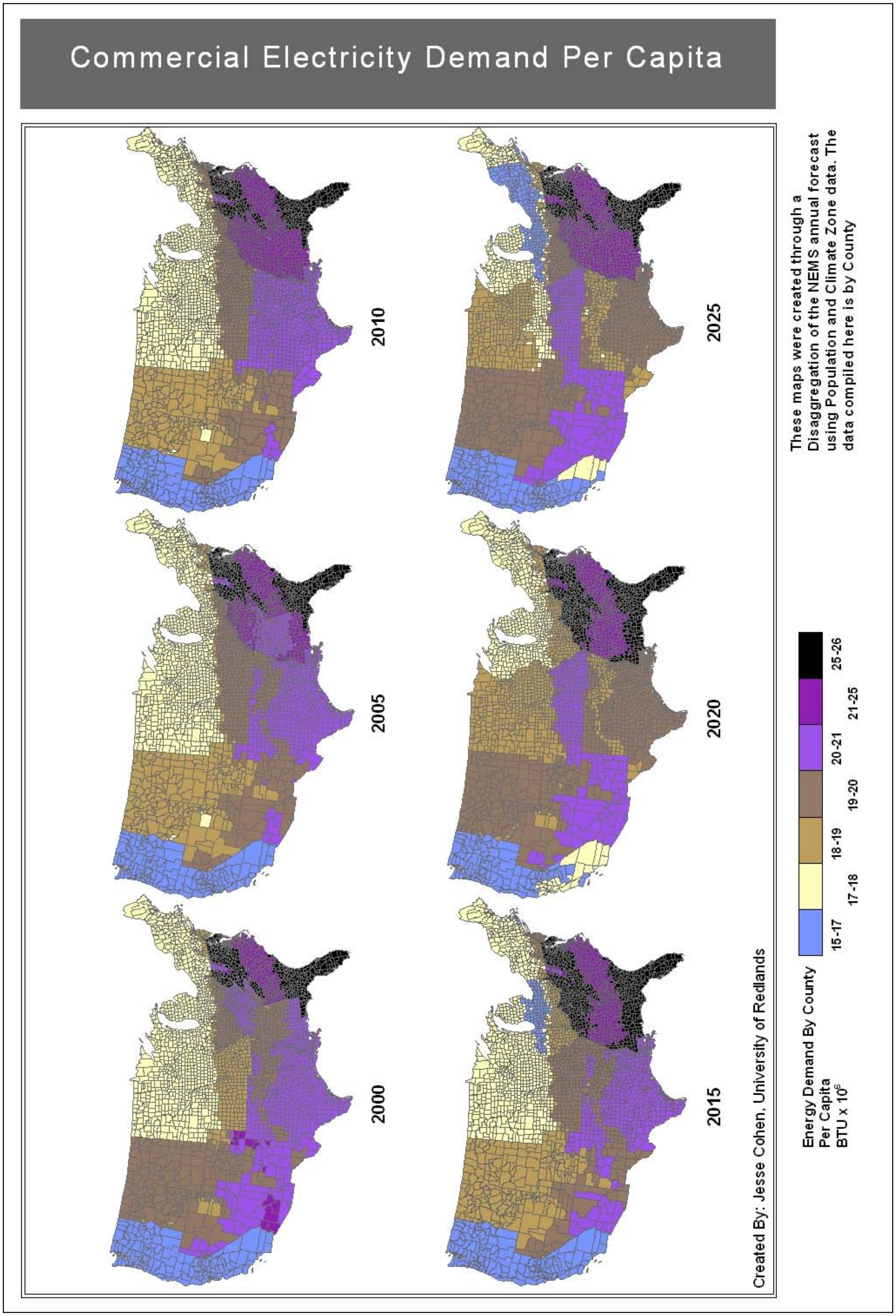
U.S. Environmental Protection Agency. 2004. "Population Projections by County." National Ambient Air Quality Standards. *http://www.epa.gov/ttn/naaqs/ozone/areas/pop/pop_proj.htm*.

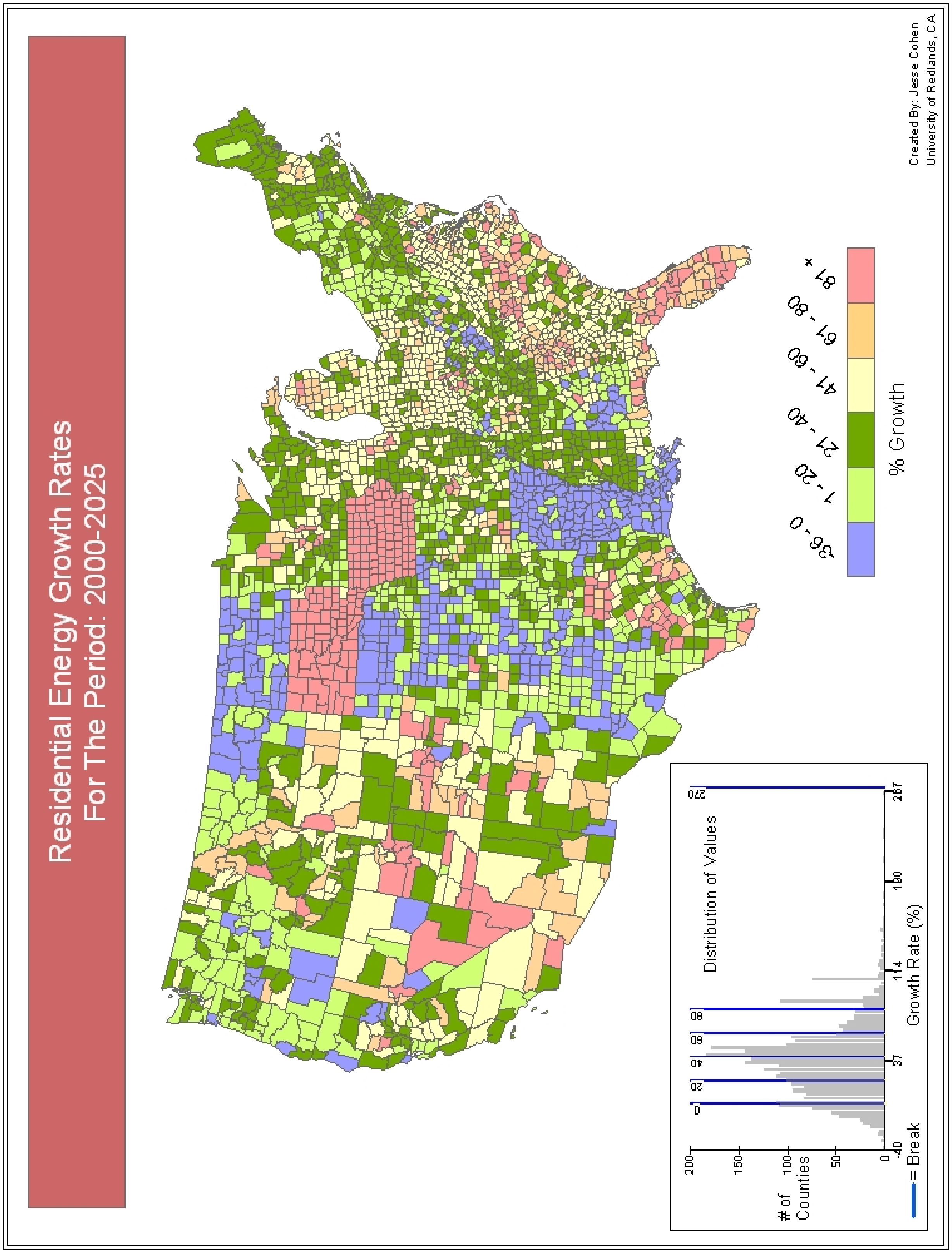
Yue, T., Y. Wang, and S. Chen. 2003. "Numerical Simulation of Population Distribution in China." *Population and Environment*, 25 (2), 141-160.

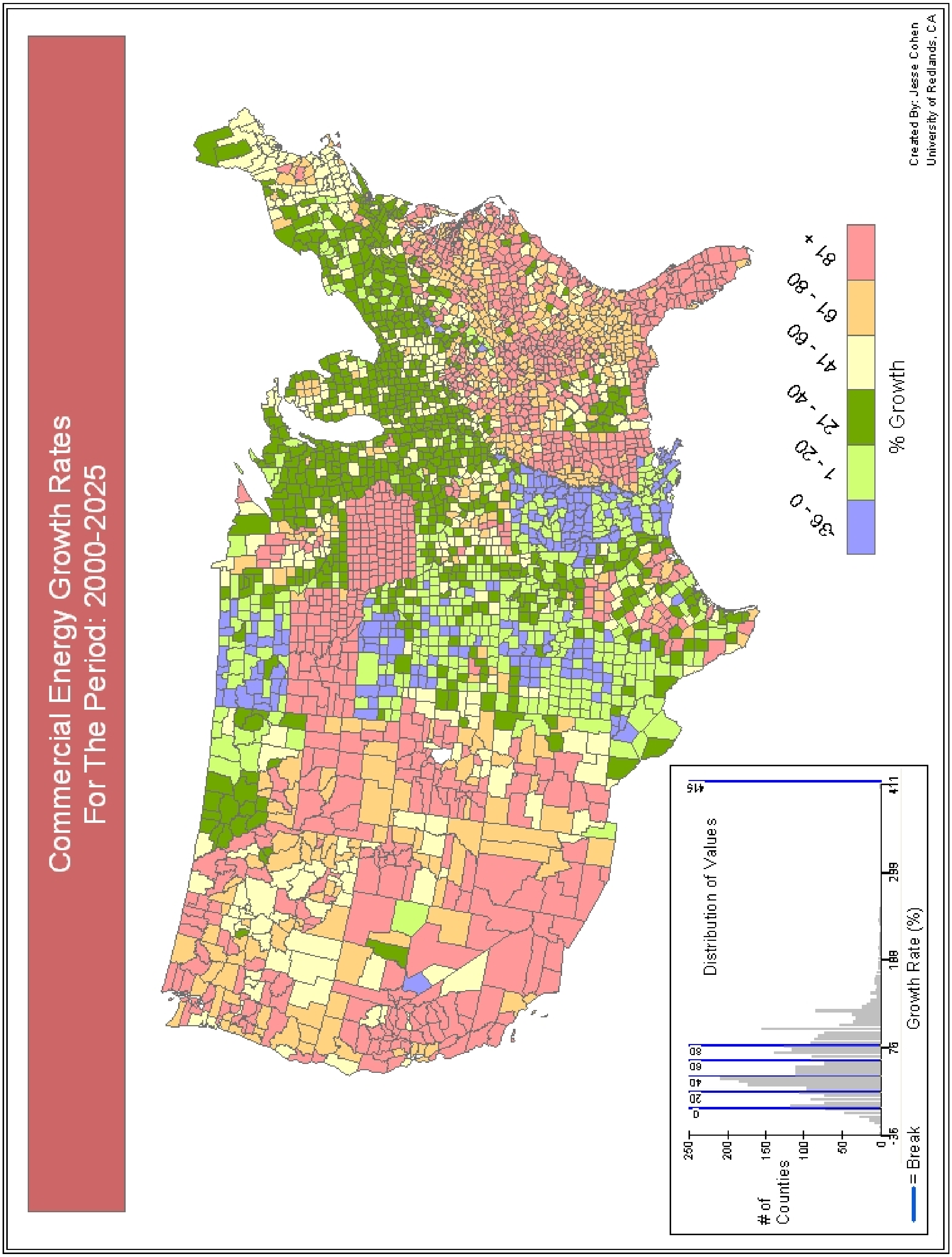
Zepp, J. 2004. "Forecasts and Simulations: Two Different Analysis Tools." Justice Research and Statistics Association. *http://www.jrsa.org/programs/forecast.htm*.

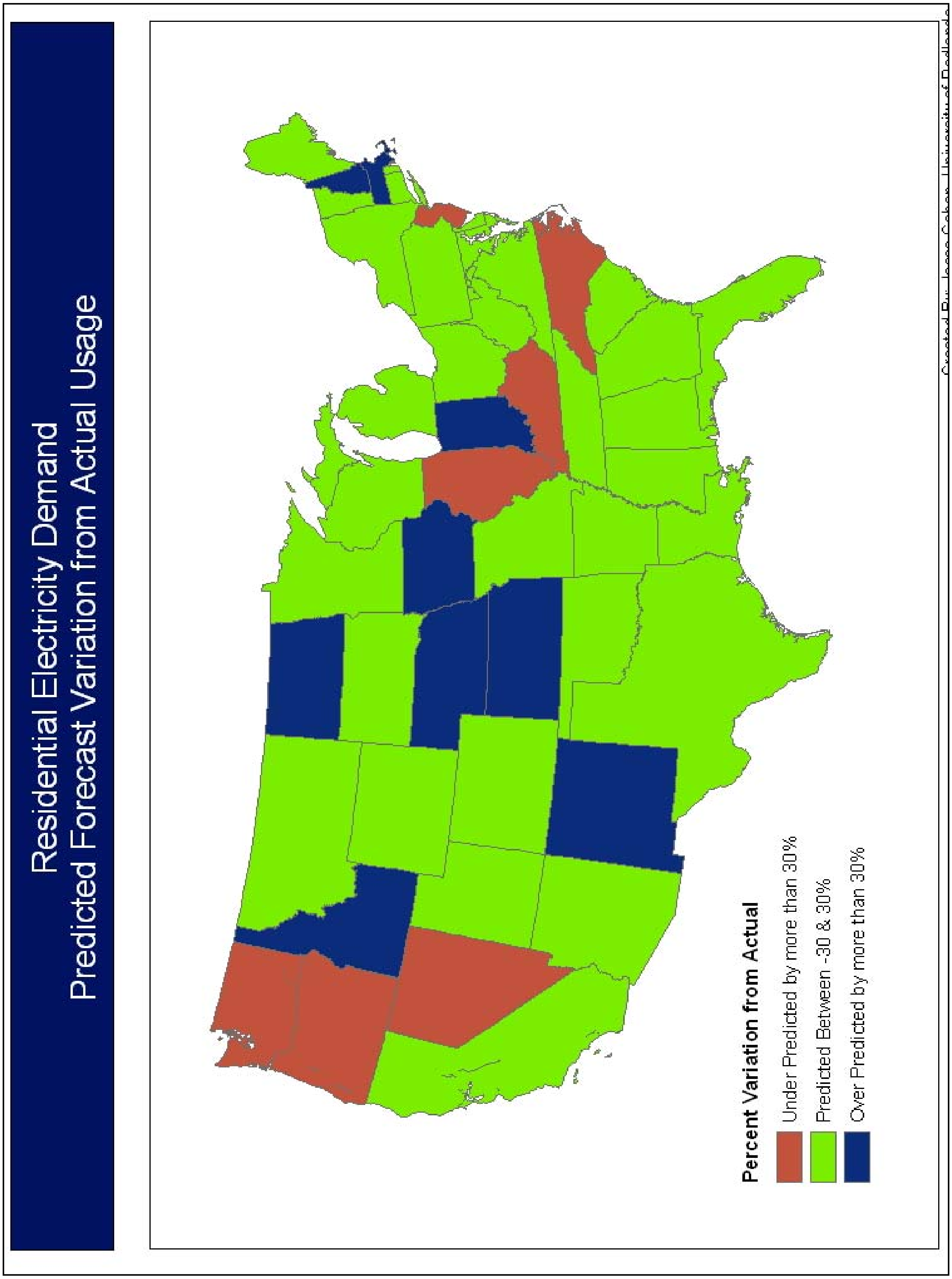
Appendix A. Map Results Gallery

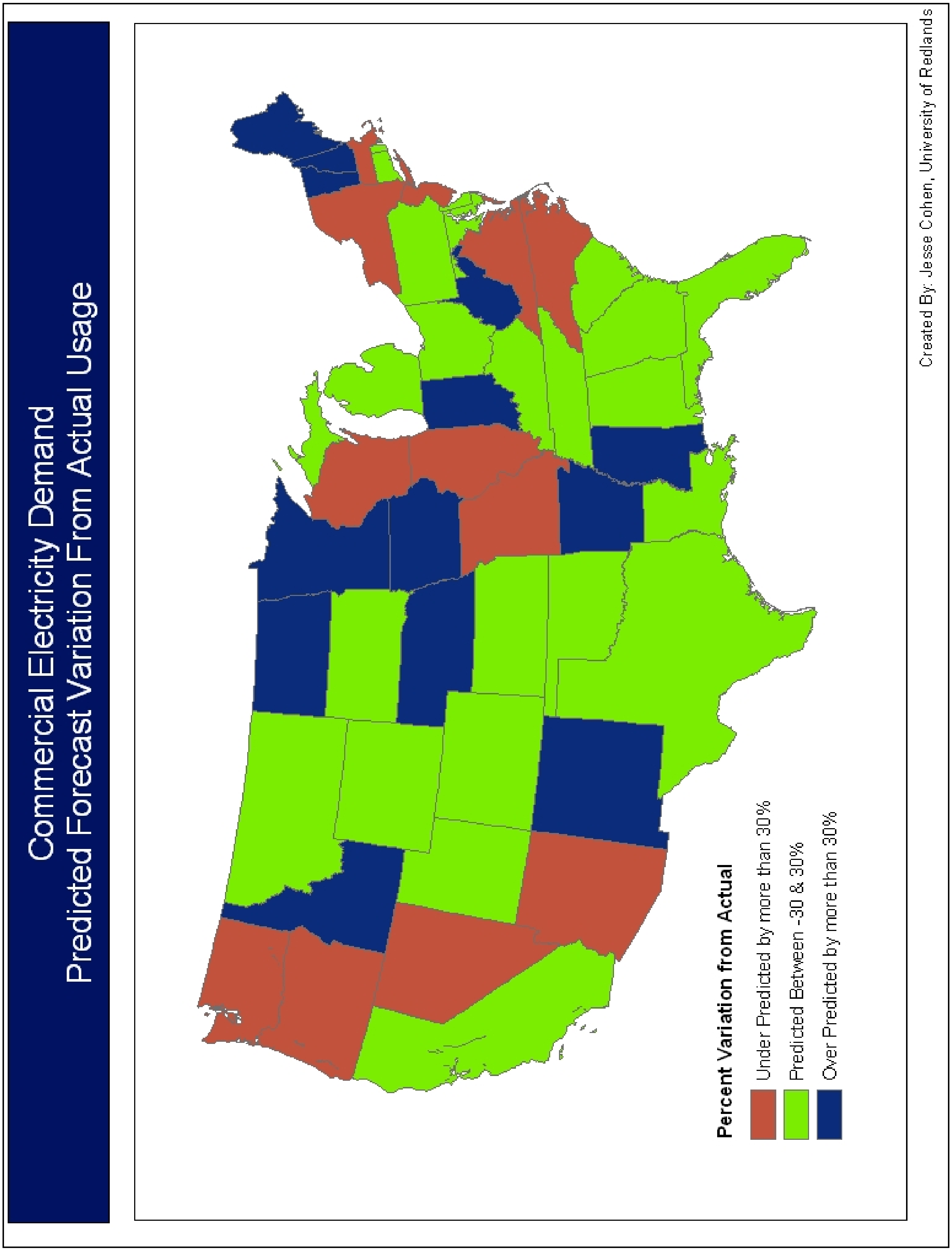

Commercial Electricity Demand
2010
2005
2000
2025
2020
2015
Created By: Jesse Cohen, University of Redlands
Energy Demand By County
BTU x 10^12
0 - 1
1-5
5 - 20
20- 40
40 - 65
65 - 100
100 - 190
These maps were created through a Disaggregation of the NEMS annual forecast using Population and Climate Zone data. The data compiled here is by County


Commercial Electricity Demand Per Capita
2000
2005
2010
2015
2020
2025
Created By: Jesse Cohen, University of Redlands
Energy Demand By County
Per Capita
BTU x 10^6
15-17
17-18
18-19
19-20
20-21
21-25
25-26
These maps were created through a
Disaggregation of the NEMS annual forecast
using Population and Climate Zone data. The
data compiled here is by County

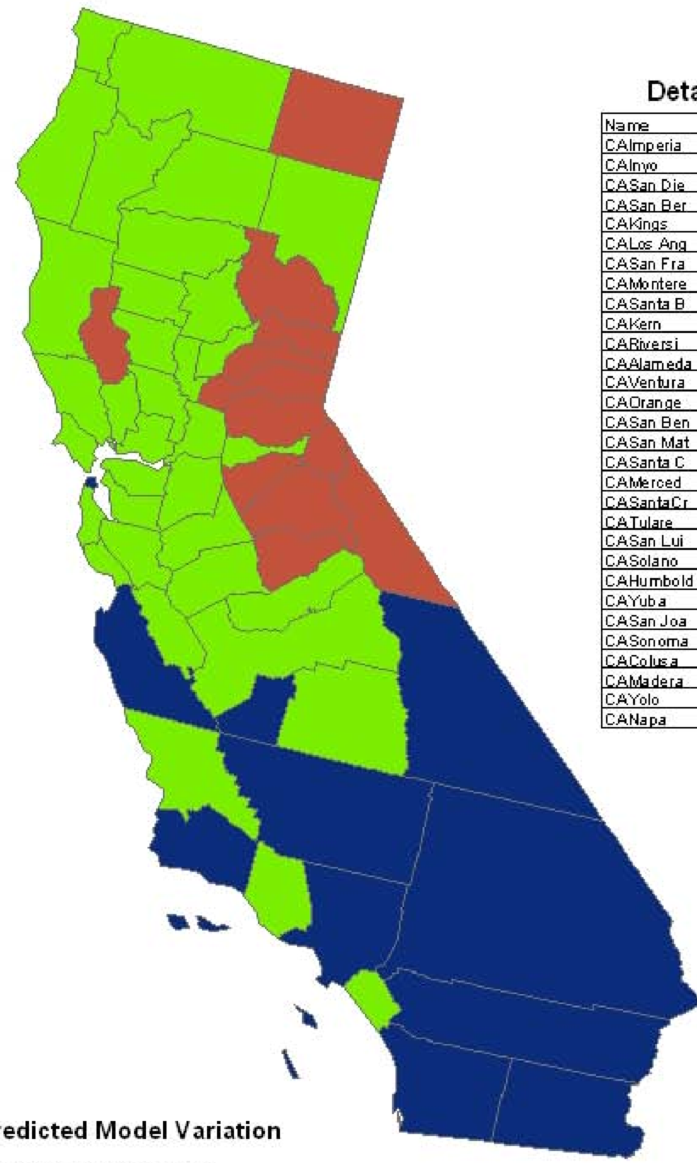

Residential Energy Growth Rates
For The Period: 2000-2025
% Growth
-36 - 0
1 - 20
21 - 40
41 - 60
61 - 80
81 +
Distribution of Values
# of Counties
Growth Rate (%)
= Break
Created By: Jesse Cohen
University of Redlands, CA


Commercial Energy Growth Rates
For The Period: 2000-2025
% Growth
-36 - 0
1 - 20
21 - 40
41 - 60
61 - 80
81 +
Distribution of Values
# of Counties
Growth Rate (%)
= Break
Created By: Jesse Cohen
University of Redlands, CA


Residential Electricity Demand
Predicted Forecast Variation from Actual Usage
Percent Variation from Actual
Under Predicted by more than 30%
Predicted Between -30 & 30%
Over Predicted by more than 30%


Commercial Electricity Demand
Predicted Forecast Variation From Actual Usage
Percent Variation from Actual
Under Predicted by more than 30%
Predicted Between -30 & 30%
Over Predicted by more than 30%
Created By: Jesse Cohen, University of Redlands

Residential Electricity Demand
Predicted Forecast Variation from Actual Usage



Details of California Counties

Name	% off actual	CAContras	8
CAImperia	-1614	CAFresno	8
CAInyo	-93	CAMarin	8
CASan Die	-78	CASutter	9
CASan Ber	-68	CAGlenn	11
CAKings	-59	CAMendoci	12
CALos Ang	-46	CAStanisl	16
CASan Fra	-45	CALassen	19
CAMontere	-44	CASacrame	23
CASanta B	-43	CAButte	24
CAKern	-41	CADel Nor	26
CARiversi	-31	CATehama	27
CAAlameda	-30	CAShasta	28
CAVentura	-30	CATrinity	28
CAOrange	-26	CAAmador	29
CASan Ben	-22	CANevada	32
CASan Mat	-22	CAPlacer	32
CASanta C	-20	CAMaripos	33
CAMerced	-14	CAEl Dora	34
CASantaCr	-13	CALake	36
CATulare	-13	CACalaver	37
CASan Lui	-10	CASierra	44
CASolano	-10	CATuolumn	46
CAHumbold	-7	CAPlumas	48
CAYuba	-7	CAAlpine	50
CASan Joa	0	CAModoc	57
CASonoma	1	CAMono	70
CAColusa	3	CASiskiyo	#DIV/0!
CAMadera	4		
CAYolo	4		
CANapa	6	Total	-26

Predicted Model Variation

% Difference from Actual

- Under Predicted More than -30 %
- Predicted Between -30 % & 30 %
- Over Predicted More than 30 %

Created By: Jesse Cohen
University of Redlands, CA

Appendix B. Disaggregation Results Spreadsheets

Residential Results by State 2000

STATE_NAME	Predicted	Actual	Difference	% off Actual
Vermont	121146320	7000000	114146320	1631
New Mexico	85860100	16800000	69060100	411
Nebraska	136539290	28500000	108039290	379
North Dakota	27042970	11600000	15442970	133
New Hampshire	25975438	12500000	13475438	108
West Virginia	67939360	33200000	34739360	105
Iowa	81549210	41000000	40549210	99
Kansas	78614190	42700000	35914190	84
Idaho	43822130	23900000	19922130	83
Arkansas	87243600	50700000	36543600	72
Indiana	163348440	97700000	65648440	67
Massachusetts	91225600	59900000	31325600	52
Mississippi	77128440	58700000	18428440	31
California	344301540	270400000	73901540	27
New York	186614420	146800000	39814420	27
Rhode Island	11551200	9100000	2451200	27
Michigan	121122640	104800000	16322640	16
Colorado	54247147	47900000	6347147	13
Delaware	13117000	12200000	917000	8
Florida	365154400	337800000	27354400	8
Minnesota	68319450	63600000	4719450	7
Maine	13284600	12700000	584600	5
Texas	413931003	398800000	15131003	4
Georgia	152401180	152000000	401180	0
South Dakota	11614560	11700000	-85440	-1
Alabama	96448200	98100000	-1651800	-2
Ohio	150411000	158600000	-8189000	-5
Louisiana	88712300	94600000	-5887700	-6
Missouri	92478250	100900000	-8421750	-8
Tennessee	115179180	125000000	-9820820	-8
Connecticut	36041000	39700000	-3659000	-9
South Carolina	78508800	86200000	-7691200	-9
Oklahoma	60051950	67000000	-6948050	-10
Maryland	67057250	81700000	-14642750	-18
Wyoming	5767400	7200000	-1432600	-20
Montana	10567347	13300000	-2732653	-21
Pennsylvania	121768230	153600000	-31831770	-21
Kentucky	44571250	79800000	-35228750	-44
Arizona	44513000	84800000	-40287000	-48
Oregon	31707880	62100000	-30392120	-49
Washington	54294900	112700000	-58405100	-52
Wisconsin	31075500	68000000	-36924500	-54
Nevada	10091640	32100000	-22008360	-69
North Carolina	25034390	158800000	-133765610	-84
New Jersey	12648200	83800000	-71151800	-85
Illinois	16680880	137000000	-120319120	-88
Virginia	6411350	128100000	-121688650	-95
Total	4072457425	4047300000	25157425	1

Commercial Results by State

2000 Millions BTU

STATE_NAME	Predicted	Actual	Difference	% off Actual
Vermont	118770980	6700000	112070980	1673
Nebraska	141234770	29800000	111434770	374
New Mexico	109907200	28600000	81307200	284
West Virginia	76792560	23400000	53392560	228
North Dakota	27341960	10200000	17141960	168
Arkansas	85809200	32300000	53509200	166
Iowa	84997970	33900000	51097970	151
Indiana	170755690	71900000	98855690	137
New Hampshire	23904944	13300000	10604944	80
Mississippi	70779280	41900000	28879280	69
Idaho	41291530	25300000	15991530	63
Kansas	58913350	44900000	14013350	31
Maine	16671400	13200000	3471400	26
California	402425020	340900000	61525020	18
Utah	34772560	29800000	4972560	17
Rhode Island	12735500	11100000	1635500	15
South Carolina	72197500	62900000	9297500	15
Michigan	139933050	125500000	14433050	12
Florida	294415600	265800000	28615600	11
Georgia	146245320	131200000	15045320	11
Minnesota	46613780	42000000	4613780	11
South Dakota	10796620	9700000	1096620	11
Pennsylvania	160917580	146700000	14217580	10
Alabama	71798600	67300000	4498600	7
Massachusetts	85612900	80000000	5612900	7
Colorado	67204493	64900000	2304493	4
New York	250267320	240300000	9967320	4
Ohio	159028800	152300000	6728800	4
Louisiana	71552300	71700000	-147700	0
Oklahoma	54524820	54600000	-75180	0
Connecticut	42043000	42600000	-557000	-1
Texas	336585237	340300000	-3714763	-1
Tennessee	89480810	91500000	-2019190	-2
Montana	13297426	14000000	-702574	-5
Maryland	80698000	90400000	-9702000	-11
Delaware	12261000	14000000	-1739000	-12
Missouri	76057260	92000000	-15942740	-17
Wyoming	7342110	10000000	-2657890	-27
Oregon	35986420	53700000	-17713580	-33
Washington	62327830	95700000	-33372170	-35
Kentucky	36505280	58900000	-22394720	-38
Arizona	42033600	82900000	-40866400	-49
Wisconsin	29164560	65000000	-35835440	-55
Nevada	9260500	24400000	-15139500	-62
North Carolina	28869760	133300000	-104430240	-78
New Jersey	16004400	114200000	-98195600	-86
Illinois	20161570	181400000	-161238430	-89
Virginia	7954370	160300000	-152345630	-95
Total	4054245730	3936700000	117545730	3

Residential Results by County

Residential 2000				
County	**Predicted**	**Actual**	**Difference**	**% off actual**
Imperial	1930000	112601	1817399	1614
Inyo	230300	119425	110875	93
San Diego	36760000	20612750	16147250	78
San Bernardino	21570000	12877424	8692576	68
Kings	1582000	992933	589067	59
Los Angeles	91090000	62585509	28504491	46
San Fran	7336000	5053382	2282618	45
Monterey	3722000	2576167	1145833	44
Santa B	3816000	2661471	1154529	43
Kern	8460000	5998546	2461454	41
Riverside	19620000	14999776	4620224	31
Alameda	13610000	10461627	3148373	30
Ventura	6982000	5384360	1597640	30
Orange	26240000	20786769	5453231	26
San Ben	480300	392396	87904	22
San Mat	6920000	5667568	1252432	22
Santa Cruz	16340000	13614447	2725553	20
Merced	1994000	1743605	250395	14
Santa Cr	2411000	2139413	271587	13
Tulare	3582000	3163056	418944	13
San Luis	2360000	2142825	217175	10
Solano	3703000	3357548	345452	10
Humboldt	1189000	1108946	80054	7
Yuba	592600	552767	39833	7
San Joaquin	5369000	5363887	5113	0
Sonoma	4254000	4292475	-38475	-1
Colusa	194300	201316	-7016	-3
Madera	1170000	1214723	-44723	-4
Yolo	1519000	1586646	-67646	-4
Napa	1177000	1248844	-71844	-6
Contras	8632000	9420924	-788924	-8
Fresno	7513000	8144783	-631783	-8
Marin	2301000	2504512	-203512	-8
Sutter	759900	832563	-72663	-9
Glenn	271400	303681	-32281	-11
Mendoc	837700	951988	-114288	-12
Stanislaus	4251000	5080679	-829679	-16
Lassen	296400	365099	-68699	-19
Sacramento	11230000	14651738	-3421738	-23
Butte	1919000	2518161	-599161	-24
Del Nor	293700	399221	-105521	-26
Tehama	524800	716550	-191750	-27
Shasta	1628000	2275899	-647899	-28
Trinity	127200	177431	-50231	-28
Amador	322800	453815	-131015	-29
Nevada	914500	1344384	-429884	-32
Placer	2297000	3398493	-1101493	-32
Mariposa	155200	232026	-76826	-33
El Dora	1539000	2330493	-791493	-34

Lake	556400	870096	-313696	-36
laver	389400	621010	-231610	-37
Sierra	28500	51182	-22682	-44
Tuolumne	529200	975873	-446673	-46
Plumas	171900	327566	-155666	-48
Alpine	10210	20473	-10263	-50
Modoc	86410	201316	-114906	-57
Mono	89790	300268	-210478	-70
Siskiyou	426000	0	-426000	No data
Total	344304910	272483424	-71821486	-26

Appendix C. Application Development

The Areal Interpolation Control Zone method offers a means to break down census division level data into county level data. The next logical step is to automate the process making it easier to run cases that represent both high and low demand scenarios. ArcGIS allows for customization of the user environment using Visual Basic for Applications (VBA). The functionality of ArcGIS is ported through a number of objects called ArcObjects. These objects allow the user to create linear or branched processes using the suite of tools available in ArcGIS.

The NEMS Disaggregation Engine Tool was intended to serve two main purposes:

1. To offer a standard and repeatable disaggregation process.
2. To allow analyst to replace the application's inputs as data sources improve and new annual results are released. The level of flexibility proposed here is crucial to keep up with annual releases of NEMS and the AEO.

User Needs Assessment

In order for the NEMS Disaggregation Engine to be usable over time, the application must be both flexible and functional for GIS and non-GIS users. Below is a short discussion of these two points.

Flexibility

The NEMS Disaggregation Engine Tool was coded to allow for change over time, as all aspects of the process evolves. Flexibility can be found in 3 main aspects of the tool:

1. *Input Datasets*: The tool requires four main input data sets before the disaggregation can run. These data sets are interchangeable with up-to-date or higher quality data.

2. *Manual "Control Zone" weighting*: The shapefile that will act as the control zone is required to have at least five classes, as in the case of the current Climate Zone shapefile. However, the tool allows for up to ten unique classes, if another control zone file is used as a substitute.

3. *Output Year(s)*: The user has the choice of which year(s) to output. The standard case will be "All Years", but the choice of individual years allows for reduced processing time and more focused analysis.

Functionality

The following levels of functionality were adhered to:

1. *User-friendly*: The NEMS Disaggregation Engine tool is made for ease of use with an intuitive GUI. The main functions to assist in ease of use are:

Easy navigation to directories and files
Comprehensive help file
Summary table of inputs

2. *Error-free*: The tool is built to run without errors. The coding requires all fields be filled with the proper data before the process is run. The main functions to assist in this process are:

Error checking of input file format
Error checking for completeness
Summary table of inputs

Graphic User Interface

The NEMS Disaggregation Engine tool (Figure 37) can be placed on any toolbar within the ArcGIS 8.0 (and above) interface. The tool is meant to be user friendly and flexible, requiring no additional training.

Figure 37: The NEMS tool as added to the ArcGIS interface

When the tool is clicked, the Graphic User Interface (GUI) opens. The GUI is preceded by a splash screen required by LBNL which details the project funding and the proprietary nature of both the data and application, along with the LBNL logo. There are two main screens to the application, the input and output interfaces.

Input Screen

The Input screen allows the user to select the main inputs of the disaggregation process (Appendix D: Figure D-3). The four main input files are the population spreadsheet (.DBF), County Shapefile, Climate Zone Shapefile, and census division electricity forecasts. Information regarding the details on formatting and necessary field is included in the Help document and can be accessed by using the “Help” button.

The next option is the scaling factor for the end-use categories (Figure D-3, D-4): residential, commercial, and unique scale. The first two categories scaling factors are hard-coded into the application. These scaling factors come from the RECS and CBECS annual reports and are discussed in more detail in Section 3.5. LBNL is also interested in exploring different scenarios of future electricity demand and because of this there is a third option “Unique Scale”. This option allows the user to provide unique weights to the control zone. This process is discussed in detail in Appendix D.

Output Screen

The Output screen sets the output file location and the years for which the disaggregation will be run (see Appendix D: Figure D-5). The year selection interface allows the user to choose anywhere from one to all of the available years to be run through the disaggregation process. The outputs are shapefiles for each year selected and by default are named "CountyPop{maxXXXX}", where XXXX is the year of interest. Since the output file cannot be changed and thus may be overwritten, it is recommended that a new directory be created for each run of the NEMS Disaggregation Engine. (Appendix D: Figure D-5)

The Summary Screen acts as a last chance to see the input and output results for the analysis (see Appendix D: Figure D-6). The screen shows the name and path of the input files, the choice of scaling factors, the years of analysis, and the output directory. The interface also gives the user a BACK button to return to the original input screen. This option keeps all of the original choices in memory until changed by the user.

Once "OK" is pressed on the summary screen, a progress bar is implemented to show the status of the application. Since there is substantial geoprocessing and up to six possible years of analysis, this feature is necessary to inform the user that the process is still running and when the tool has completed the procedures.

Development Issues

The NEMS Disaggregation Engine was developed using the ArcObjects suite of objects. During development there were a few programming problems encountered. Solutions for these problems are detailed in Appendix F.

Summary

ArcObjects is a complex set of objects with a steep learning curve. Although the potential for complex functionality exists, it would be too time consuming and impractical at present to access many of these features. Therefore, due to the limitations of time, some aspects of the functionality had to be sacrificed. Despite this, the application was crafted to fit the majority of user needs, providing the data necessary to reliably complete analyses.

There are many possible improvements to the coding behind the application. Fortunately, the application was coded in a modular fashion. This creates the opportunity to insert more complex blocks of code, which enhances the application's functionality over time. Some of these are discussed in Chapter 7.3.

Appendix D. NEMS Disaggregation Engine Users Guide

The purpose of the Users Guide is to enable an individual to work with the tool and create usable results in a timely fashion.

To begin, open the ArcMap template "NEMS" which contains the tool and its VBA code. The tool must be placed on a toolbar within the ArcMap interface before use. See ArcGIS help for instructions on adding items to a toolbar.

Figure D-1: The NEMS tool button

Once the button is pressed the splash screen (Figure A-2) will come up containing information on the project funding and the proprietary nature of its use. This document can be read in more detail from the "Input data" screen by pressing the "About" button.

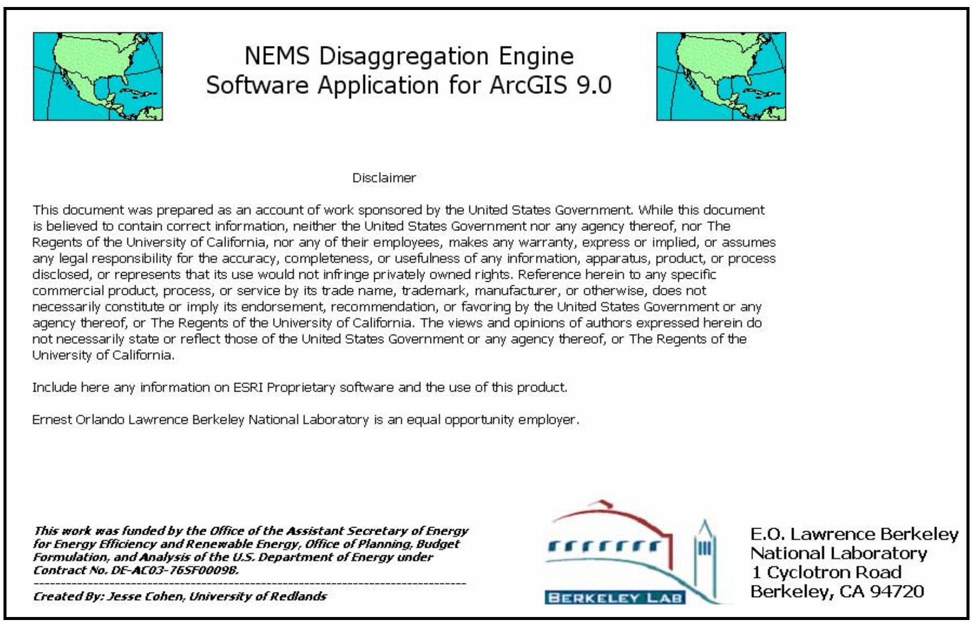


Figure D-2: The Splash Screen

Once the splash screen disappears and the "Input Data" screen appears (Figure A-3), begin using the tool by following the steps below.

1. The first step is to search for each of the necessary input files by pressing the button found on the side of each textbox.

2. After selecting each of the necessary input files a Scaling Factor is required. The first two radio buttons refer to hard-coded scaling factors that can be found in the HELP file. If either the "Commercial" or "Residential" button is pressed proceed to Step 5. OR…

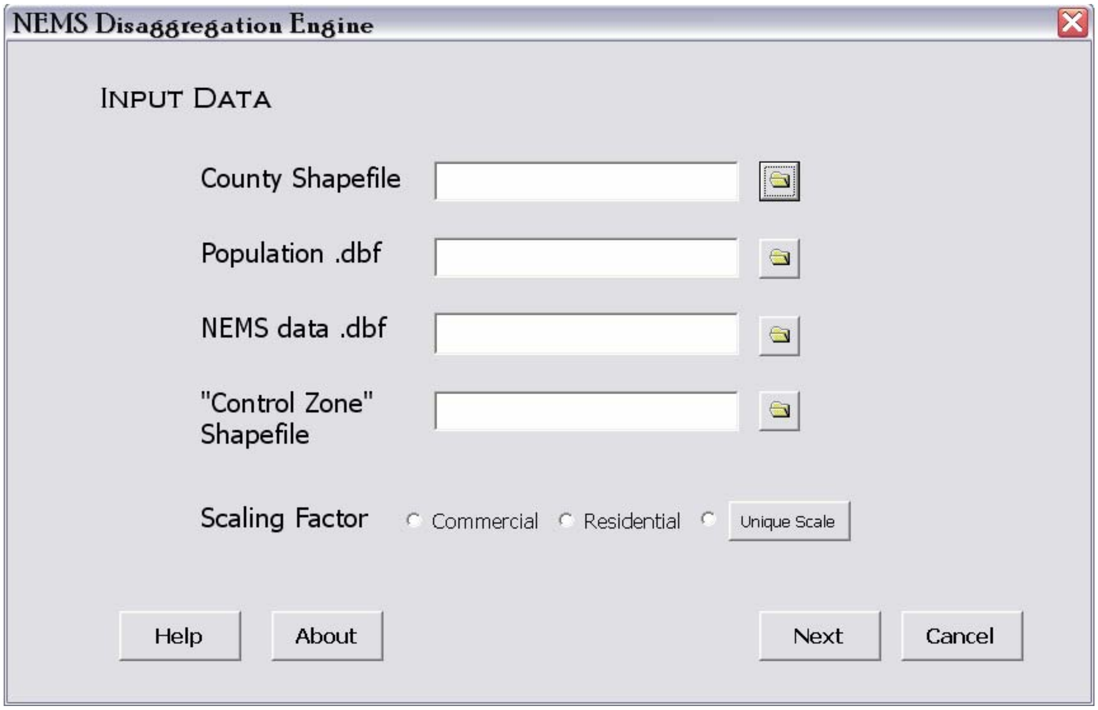


Figure D-3: The Input Data Screen

3. Press the Unique Scale button to bring up the Unique Scale screen (Figure A-4). This screen offers a chance to test various scenarios or work with control zone data with 6-10 classes. A minimum of 5 weights must be entered to continue.

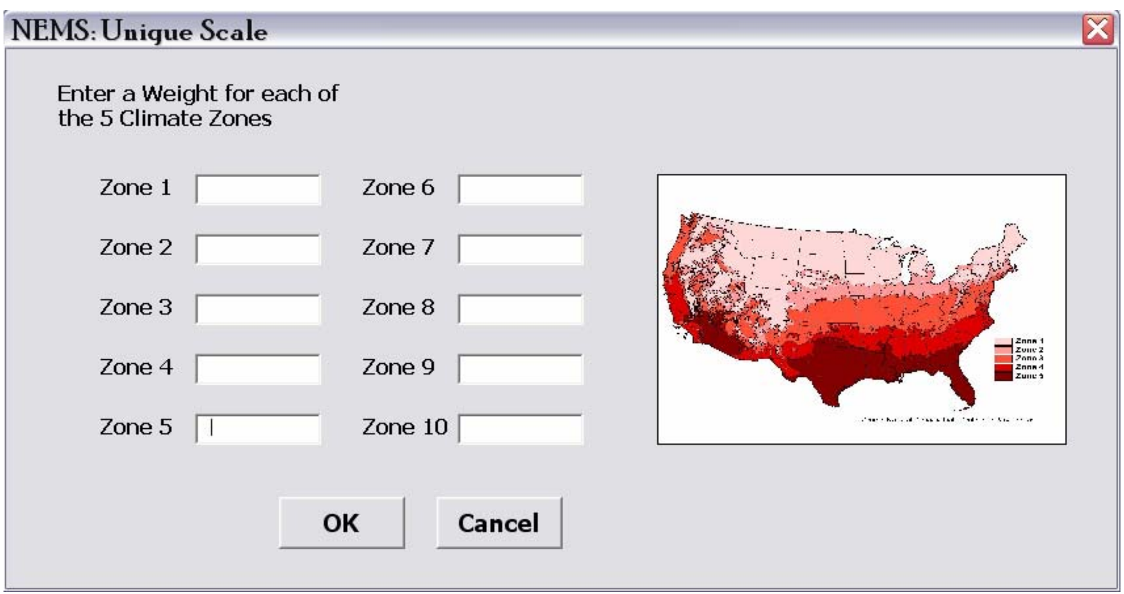


Figure D-4: The Unique Scale screen

Press **OK** when finished.

Once the Input files are located and the weighting factors are set, **Press Next**.

The Output Data screen appears (Figure A-5).

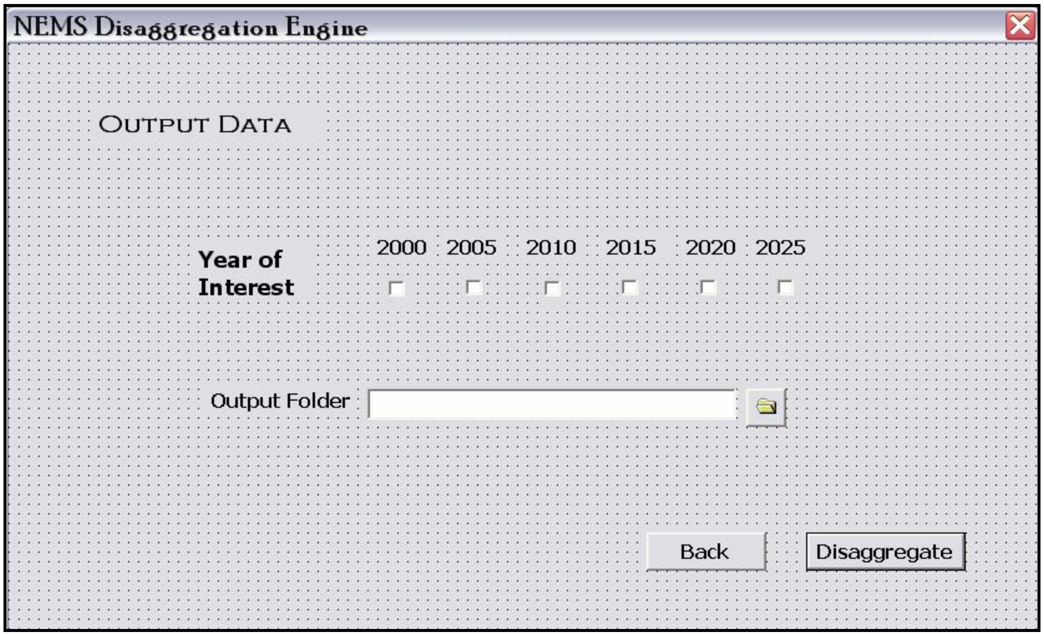


Figure D-5: The Output Data screen

4. **Select** as many years as necessary from the "Year of Interest" checkboxes.

Press the Output Folder button to choose Output file directory.

To go back to the first screen **Press the BACK button.**

Press **Disaggregate** to continue.

Read the Summary screen to assure the proper inputs and outputs.

5. **Press Change** to return to Input screen OR……

Press OK to begin processing.

A pop-up window saying "Done" (Figure A-6) will appear when the process is complete. **Press OK.**

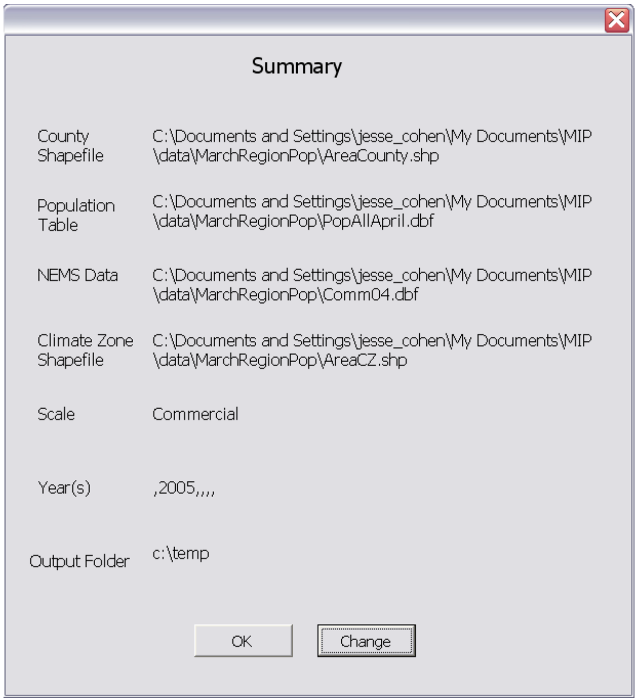


Figure D-6: The Summary Screen

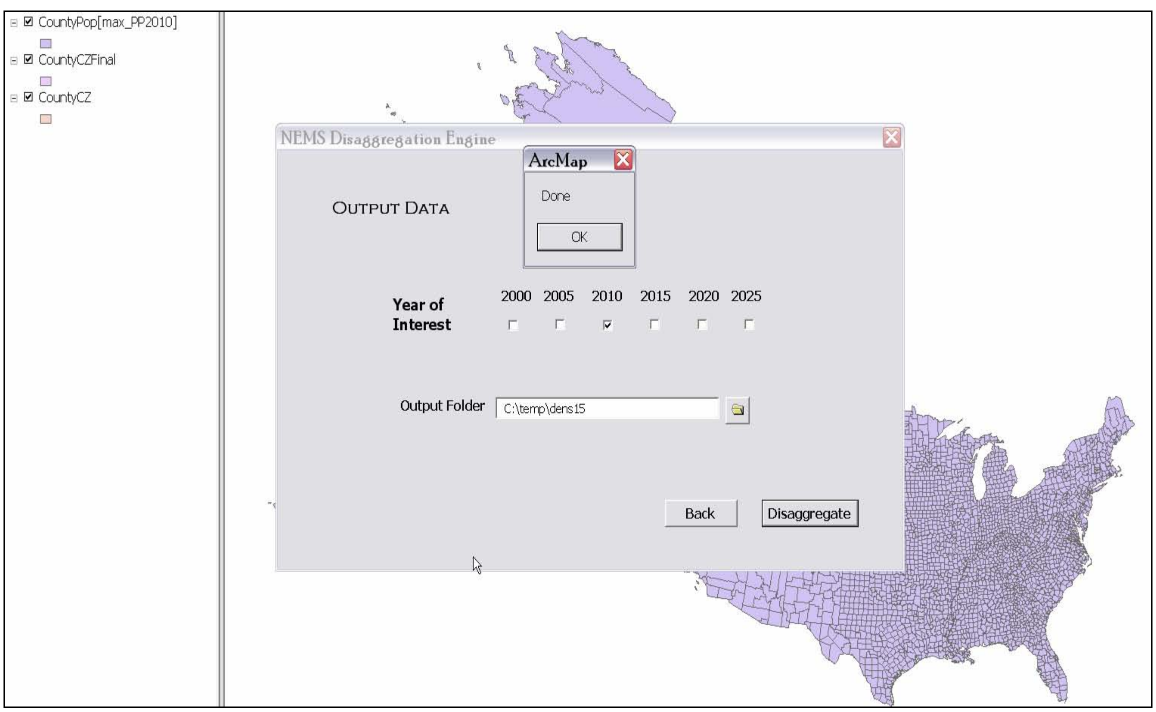


Figure D-7: A snapshot of the completed disaggregation

Appendix E. Development Issues

During the process of developing the NEMS Disaggregation Engine there were some issues and work-arounds that merited a more detailed discussion. Namely these three issues:

- Joining two tables to a shapefile,
- returning summary statistic for a field in a table, and
- looping through calculations using the field calculator.

There were two major data tables that had to be joined to a single shapefile. There are several ways to do this in ArcObjects. The table could be joined, then saved as a new shapefile, or joined and not saved as a new shapefile. The problem with saving a new shapefile is that there will be an extra file to be deleted. The problem was that two tables cannot programmatically be joined to the same shapefile. The process would occur but ArcGIS would only see the last join and ignore the first. To work around this issue, the first table was joined to the shapefile and the second table was only related to the shapefile-. The solution worked, but the look-up time on a related table was substantially longer than that of a joined table. Unfortunately, this work-around markedly cost the application in performance.

Calculating the sum for a column in an attribute table was necessary for some of the field calculations. Due to the ArcObjects' limitations, an extra step had to be taken which programmatically looped through each row in the field to add each value to the next, eventually ending up with the sum of the entire field. This work-around did not heavily affect the application's performance.

In the final stages of the disaggregation model, the application needed to loop through a table and separate out all the rows according to census regions. After that was done, the loop was supposed to continued and perform a number of calculations. The problem occurred when the census regions could not be distinguished in order to separate the data. Each time a new name was found, the program was supposed to save it in an array and continue until it found the next new name. This piece of code did not work properly. To work around this problem, the 9 census region names were entered manually into the array prior to the other operations by the program.

The work-arounds needed to get past these issues took a toll on the performance of the application. From personal assessment it appeared that the application could be running up to 2x faster. With more time and experience, these bugs could be worked through and solved for Version 2-

Appendix F. NEMS Disaggregation Engine Help File

Created by: Jesse Cohen, University of Redlands
Created on: 5/12/04
Purpose: To automate the disaggregation process for NEMS data, from the census division to the census county level.

This file can be found by pressing the "Help" button on the Input Data screen (Figure A-5)

Input Details:

1. County shapefile
   a. Must have field named "FIPS" – A five digit code referring to county and state
   b. Must have field named "Area2" – A field containing the area of the county polygons (Albers Equal Area projection)

2. Population table

3. NEMS table: By census division (spelled as below)
   a. NE
   b. SAtlantic
   c. ENCentral
   d. ESCentral
   e. WSCentral
   f. WNCentral
   g. Mountain
   h. Pacific
   i. MidAtlantic

4. Climate Zone shapefile
   a. Attribute table must have a field with climate zone ID's as an integer in a field named "ZONE_NUM".
   b. The file must have a minimum of 5 zones and a maximum of 10 that are sequentially numbered starting from 1.

***** Important*****

The output folder location must be changed each time the tool is run. Without this errors will occur because of redundant file names!!! Because of this it is also recommended that file names are RENAMED after each running of the tool.

ALSO--- All field names are case sensitive and must be spelled correctly as above!!!

Output Details:

Output Files: All fields are Double types (meaning they can take decimals and very large numbers)

1. CountyCZ: Combines- ClimateZone, County shapes, and Population table and fields
   a. Area1 – new area after County and CZ have been intersected
   b. PercArea – Area1 (control zone / original area for the County (Area2)
   c. CZ area- The PercArea * Zone_NUM_1

2. CountyCZFinal: Dissolves the CountyCZ based on the UniqueID
   a. Sum_CZArea – adds all of the CZ area for each county together to get a total CZ value for each county between 1-5. Can be in decimal form
   b. Creates new fields before copy by year occurs

3. Each Year will be output as a Shapefile with the following attributes:
   a. Influence – Classified CZ value/county according to residential/commercial/unique scale * Population for year at hand
   b. PercInflu – Influence / sum of influence per census division
   c. Energy_Cnty – PercInflu * NEMS energy demand per division per year
   d. Name: CountyPop{maxXXXX} with XXXX being the year

Appendix G. Glossary

AEO: The Annual Energy Outlook, released by the Department Of Energy

Btu (British Thermal Unit): A unit of energy consumed by or delivered to a building. A Btu is defined as the amount of energy required to increase the temperature of 1 pound of water by 1 degree Fahrenheit, at normal atmospheric pressure. Energy consumption is expressed in Btu to allow for consumption comparisons among fuels that are measured in different units.

CBECS: Commercial Building Energy Consumption Survey.

Census Region and Division: A geographic area consisting of several States defined by the U.S. Department of Commerce, Bureau of the Census

Climate Zone: One of five climatically distinct areas, defined by long-term weather conditions which affect the heating and cooling loads in buildings. The zones were determined according to the 45-year average (1931-1975) of the annual heating and cooling degree-days (base 65 degrees Fahrenheit). An individual building was assigned to a climate zone according to the 45-year average annual degree-days.

Commercial Building: A building with more than 50 percent of its floor space used for commercial activities. Commercial buildings include, but are not limited to, the following: stores, offices, schools, churches, gymnasiums, libraries, museums, hospitals, clinics, warehouses, and jails. Government buildings were included except for buildings on sites with restricted access, such as some military bases. Agricultural buildings, parking garages, residences, and manufacturing/industrial buildings are excluded. In 1995, commercial buildings on manufacturing sites were also excluded.

Cooling Degree-Days (CDD): A measure of how hot a location was over a period of time, relative to a base temperature. In this report, the base temperature is 65 degrees Fahrenheit, and the period of time is one year. The cooling degree-day is the difference between that day's average temperature and 65 degrees if the daily average is greater than 65; it is zero if the daily average temperature is less than or equal to 65. Cooling degree-days for a year are the sum of the daily cooling degree-days for that year.

Degree-Days 45-Year Average: The average of the total annual heating and cooling degree-days (base, 65 Degrees Fahrenheit) in each NOAA Division, for the 45 years, 1931 through 1975. Computed from the Division's daily temperature averages for each year in question and used to assign individual buildings to climate zones.

Heating Degree-Days (HDD): A measure of how cold a location was over a period of time, relative to a base temperature. In this report, the base temperature used is 65 degrees Fahrenheit, and the period of time is one year. The heating degree-day is the difference between that day's average temperature and 65 degrees if the daily average is less than 65; it is zero if the daily average temperature is greater than or equal to 65. Heating degree-days for a year are the sum of the daily heating degree-days for days that year.

ArcIMS: Internet Map Server, a ESRI product capable of dynamically serving maps and data on the World Wide Web

Layer: A file that stores symbology and displays information for a given vector or raster data set. The layer doesn't actually store the data, but points to its physical location.

Natural Breaks: A data classification method that partitions data into classes using an algorithm, often called Jenks' optimization. The algorithm calculates groupings of data values and, possibly, a number of classes based on the smallest possible total error (the sum of absolute deviations about the class median or, alternatively, the sum of squared deviations about the class mean).

NEMS: National Energy Modeling System

Object: In object-oriented programming, an instance of the data structure and behavior defined by a class.

Raster: A spatial data model made of rows and columns of cells. Each cell contains an attribute value and location coordinates; the coordinates are contained in the ordering of the matrix, unlike vector structure which stores coordinates explicitly. Groups of cells that share the same value represent geographic feature.

RECS: Residential Energy Consumption Survey

Residential: Activities related to use as a dwelling for one or more households. Buildings that contain commercial activities but had 50 percent or more of their floor space devoted to residential activities are considered out-of-scope.

Shapefile: A vector data storage format for storing the location, shape, and attributes of geographic features. A shapefile is stored in a set of related files and contains one feature class

Uncertainty: Uncertainty arises from the problematic attempt to represent real-world objects, which are often complex and indefinite, in a digital database. It is the degree to which one is unsure that the data provides a good picture of the part of the world being portrayed.

Vector: A coordinate-based data model that represents geographic features as points, lines, and polygons. Each point feature is represented as a single coordinate pair, while line and polygon features are represented as ordered lists of vertices. Attributes are associated with each feature, as opposed to a raster data model, which associates attributes with grid cells.

Visual Basic – Microsoft Programming Language