

ABSTRACT

HONG, TAO. Long-Term Spatial Load Forecasting Using Human-Machine Co-construct Intelligence Framework. (Under the direction of Dr. Simon Hsiang).

This thesis presents a formal study of the long-term spatial load forecasting problem: given small area based electric load history of the service territory, current and future land use information, return forecast load of the next 20 years. A hierarchical S-curve trending method is developed to conduct the basic forecast. Due to uncertainties of the electric load data, the results from the computerized program may conflict with the nature of the load growth. Sometimes, the computerized program is not aware of the local development because the land use data lacks such information. A human-machine co-construct intelligence framework is proposed to improve the robustness and reasonability of the purely computerized load forecasting program. The proposed algorithm has been implemented and applied to several utility companies to forecast the long-term electric load growth in the service territory and to get satisfying results.

Long-Term Spatial Load Forecasting
Using Human-Machine Co-construct Intelligence Framework

by
Tao Hong

A thesis submitted to the Graduate Faculty of
North Carolina State University
In partial fulfillment of the
Requirements for the degree of
Master of Science

Operations Research
and
Industrial & System Engineering

Raleigh, North Carolina

Oct. 2008

APPROVED BY:

Simon Hsiang
Committee Chair

Shu-Cherng Fang

Yahya Fathi

DEDICATION

To my family.

BIOGRAPHY

Tao Hong was born in 1983 in China. He received his Bachelor of Engineering degree in Control Science and Engineering from Tsinghua University, Beijing in July 2005, and his Master of Science degree in Electrical Engineering from North Carolina State University in May 2008. He is now pursuing his second Master of Science degree with co-majors in Operation Research and Industrial & Systems Engineering at North Carolina State University.

ACKNOWLEDGEMENTS

I would like to thank my mother, Ms. Baohui Yin, and father, Mr. Jinmin Hong, for their continuous support, mentally and financially, for my studies abroad. I would like to thank my fiancée, Ms. Pu Wang, for her love, encouragement and tolerance.

I would like to thank my advisor Dr. Simon Hsiang, for his precious time and valuable advice during the past two years. Dr. Hsiang's enthusiasm for research and diligent working attitude has been a great example to me. I would like to thank Dr. Yahya Fathi, for accepting my application to the Operations Research program, for helping me with the tedious paper work required for my enrollment, and for serving on my committee. I would like to thank Dr. Shu-Cherng Fang for serving on my committee and for offering me a cubical in his group so that I didn't feel lonely or desperate while I'm working on my thesis. I would like to thank my former advisor, Dr. Mo-Yuen Chow for helping me to develop the ability of independent thinking.

I would like to thank my former roommate, former lab-mate, and present officemate, Dr. Le Xu, for picking me up from RDU airport when I first came to Raleigh, helping me settle down during the first year at NC State, for sharing with me his insights on research, and for guiding me to complete my first spatial load forecasting project, which had lead to this thesis.

I would like to thank my manager, Dr. Richard Brown, for the guidance in my career planning and financial support during the past year. I would like to thank my mentor, Mr.

James Burke, for advising me on the principles of being an engineer, for teaching me the fundamentals of power distribution systems, and for helping build my expertise in this field. I would like to thank my colleagues, Mr. Lee Willis, Mr. Philips Edmunds, and Ms. Jia Wang, from Quanta Technology for the support and suggestions they gave during the load forecasting projects.

I would like to thank all of my friends at NC State University for the memorable time we have had together.

TABLE OF CONTENTS

LIST OF TABLES.....	viii
LIST OF FIGURES.....	ix
1 INTRODUCTION.....	1
1.1 OVERVIEW OF SPATIAL LOAD FORECASTING.....	1
1.2 DECISION MAKING PROCESS.....	5
1.3 LITERATURE REVIEW.....	7
1.4 ORGANIZATION OF THE THESIS.....	12
2 HIERARCHICAL TRENDING METHOD.....	13
2.1 OVERALL PROGRAM STRUCTURE.....	13
2.2 DATA OUTLOOK.....	16
2.2.1 <i>Historical Load</i>	16
2.2.2 <i>Weather Normalization</i>	17
2.2.3 <i>Horizon Year Load</i>	19
2.3 DATA PREPROCESSING.....	21
2.3.1 <i>Small Area Based Historical Load</i>	21
2.3.2 <i>Service Territory Map</i>	22
2.3.3 <i>Weather Normalization</i>	22
2.4 DETERMINING HORIZON YEAR LOAD USING LAND USE INFORMATION.....	25
2.5 NEIGHBORHOOD MODULE.....	29
2.6 BOTTOM-UP AGGREGATION: S-CURVE PARAMETER TUNING.....	32
2.6.1 <i>Format of S-curve</i>	32
2.6.2 <i>Problem formulation</i>	36

2.7	TOP-DOWN ALLOCATION: MULTI-OBJECTIVE OPTIMIZATION.....	36
3	HUMAN-MACHINE CO-CONSTRUCT INTELLIGENCE	39
3.1	MOTIVATION.....	39
3.2	DETERMINING LOAD DENSITIES USING GREEDY HMCCI.....	42
3.2.1	<i>Greedy Selection Procedure</i>	42
3.2.2	<i>Greedy Critique Procedure</i>	45
3.3	CUSTOMIZED SCENARIOS.....	48
3.3.1	<i>Revising Horizon Year Load</i>	48
3.3.2	<i>Adding New Business Information</i>	49
4	IMPLEMENTATION AND RESULTS.....	51
4.1	INITIALIZATION.....	51
4.2	WEATHER NORMALIZATION.....	54
4.3	HORIZON YEAR LOAD SOLVER.....	56
4.4	SERVICE TERRITORY MAP.....	57
4.5	ENGINE.....	61
4.5.1	<i>Bottom-Up Aggregation</i>	61
4.5.2	<i>Top-Down Allocation</i>	62
4.6	FORECAST RESULTS.....	63
4.6.1	<i>Result Display Setup</i>	63
4.6.2	<i>Final Forecast Results</i>	64
5	CONCLUSION AND FUTURE WORK	68
5.1	CONCLUSION.....	68
5.2	FUTURE WORK.....	70
	REFERENCES	71

LIST OF TABLES

Table 3.1 Load densities for selection procedure	42
Table 3.2 Using critique procedure to calculate load densities	47
Table 4.1 Computer run time under different forecast speed settings	59

LIST OF FIGURES

Figure 1.1 Map of Madison, WI	2
Figure 1.2 Electric load distribution of Madison, WI.....	3
Figure 1.3 Load growth distribution of Madison, WI.....	3
Figure 1.4 Bottom-up aggregation & top-down allocation.....	11
Figure 2.1 An automated computer program structure of hierarchical trending method	13
Figure 2.2 Typical data format of transformer based historical load.....	17
Figure 2.3 Typical data format for weather normalization	18
Figure 2.4 Land use code	19
Figure 2.5 Current land use data	20
Figure 2.6 Future land use data.....	20
Figure 2.7 Small area based historical load data.....	21
Figure 2.8 Service territory map	22
Figure 2.9 Weather normalized load vs. actual load.....	23
Figure 2.10 Effect of HYL on the overall forecast (1)	26
Figure 2.11 Effect of HYL on the overall forecast (2)	26
Figure 2.12 Neighborhood table	29
Figure 2.13 Potential amount of sub-regions in each level.....	31
Figure 2.14 Bottom-up aggregation.....	33
Figure 2.15 S-curve shapes affected by HYL parameter	34
Figure 2.16 S-curve shapes affected by slope parameter.....	35

Figure 2.17 S-curve shapes affected by ramp up time parameter	35
Figure 2.18 Top-down allocation.....	37
Figure 3.1 A semi-automated hierarchical trending method with HMCCI framework.....	41
Figure 4.1 “Initialization” worksheet.....	52
Figure 4.2 “Weather Normalization” worksheet	54
Figure 4.3 “All Data” worksheet – weather normalized historical load	55
Figure 4.4 “HYL Solver” worksheet	56
Figure 4.5 “Service Territory” worksheet.....	58
Figure 4.6 “Service Territory” worksheet – information board.....	58
Figure 4.7 Temporary results of bottom-up aggregation	61
Figure 4.8 “Bottom-Up Solver” worksheet	62
Figure 4.9 “Top-Down Solver” worksheet	62
Figure 4.10 “Result Display Setup” worksheet	63
Figure 4.11 “Forecast Incremental Load” worksheet	64
Figure 4.12 “Forecast Actual Load” worksheet.....	64
Figure 4.13 “Forecast Map” worksheet	65
Figure 4.14 Forecast map selection	66
Figure 4.15 Forecast incremental load map (2010)	66
Figure 4.16 Forecast actual load map (2010).....	67

1 Introduction

1.1 Overview of Spatial Load Forecasting

If quality is defined by what the customers want, one way to measure the success of an electric power distribution system is to deliver reliable electric power to customers spread throughout utility's service territory. Figure 1.1 shows a Google map of Madison in Wisconsin, and Figure 1.2 shows its electric load distribution, which contains a pattern that is typical of a medium-sized city in the United States.

- 1) There is high load density in and around the central business district (downtown) area, while a lower density in the outlying suburban areas.
- 2) Even in the suburbs, the load density along the major industrial areas can be higher than that of the downtown areas which are filled with residential customers and office parks.
- 3) In rural areas, load density is very low because homes, commercial areas, and industrial areas are spread far apart.
- 4) Due to the use of irrigation pumps and oil pumps in petroleum fields, the load density can be surprisingly high in comparison with other rural areas.

In the planning stage, a distribution planner needs to provide an economic expansion plan to meet the future load growth in the service territory based on an incremental load distribution map (Figure 1.3). Such a map of density, namely a spatial load forecast [25], must provide temporal, spatial, and magnitude information of load growth in sufficient detail, and with

sufficient accuracy to allow for quality planning of electric power transmission and distribution facilities. Assessing the impacts of alternatives for conjunctive areas also requires attention to processes that take place on different spatial and temporal scales. Information scale is the spatial and temporal scale of the information required. Generally, a strategic resources manager (for example the local, regional or national government) needs information on a scale relative to their responsibilities and authorities. This level of information is likely to differ from the level desired by operational managers dealing with day-to-day issues. Information at scales smaller than what is needed is seen as being 'noisy'. Information at scales larger than what is needed is not relevant or helpful. Specifically, in load forecasting, the scale, or resolution, of the map is also depends on the computing capability to generate historical load data.

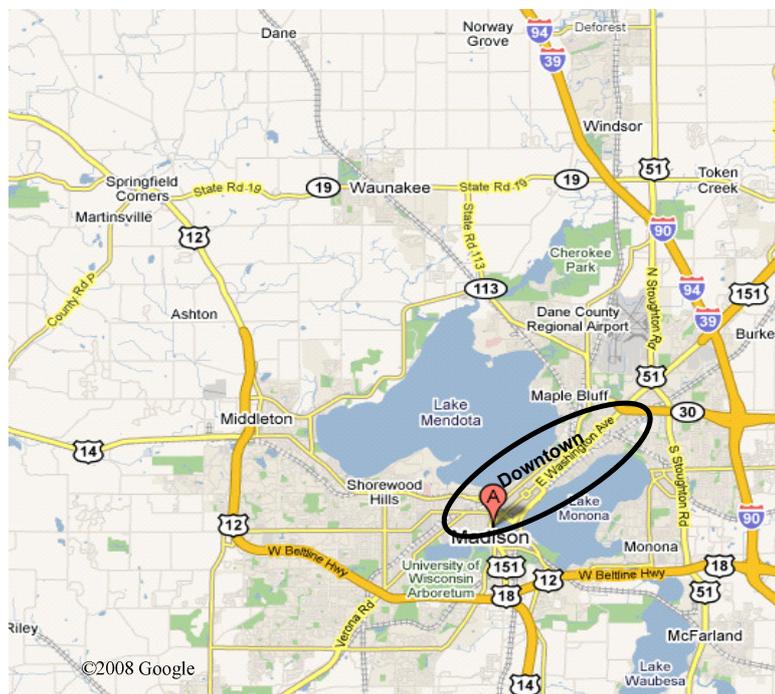


Figure 1.1 Map of Madison, WI

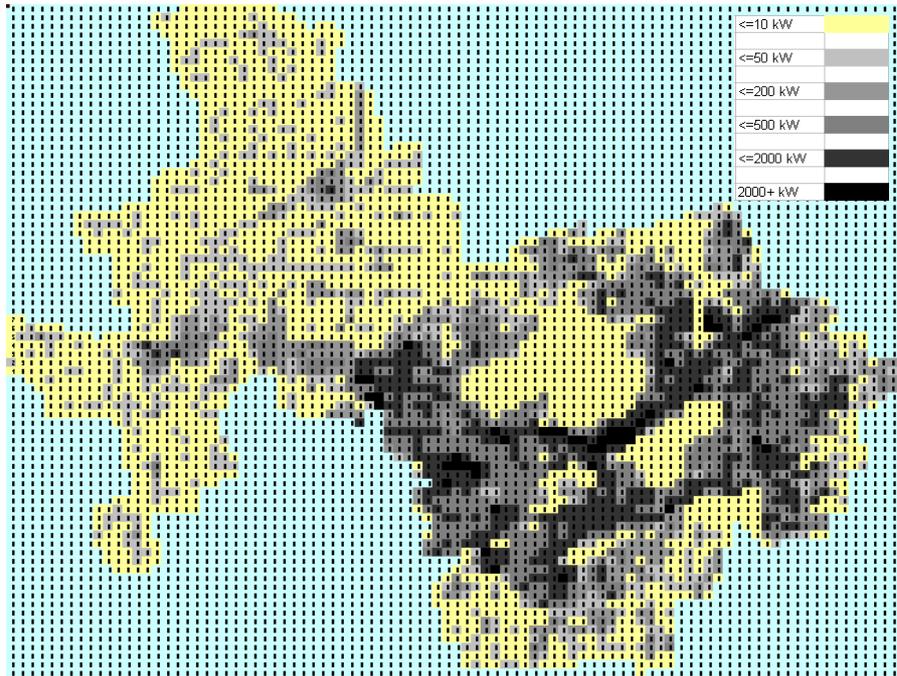


Figure 1.2 Electric load distribution of Madison, WI

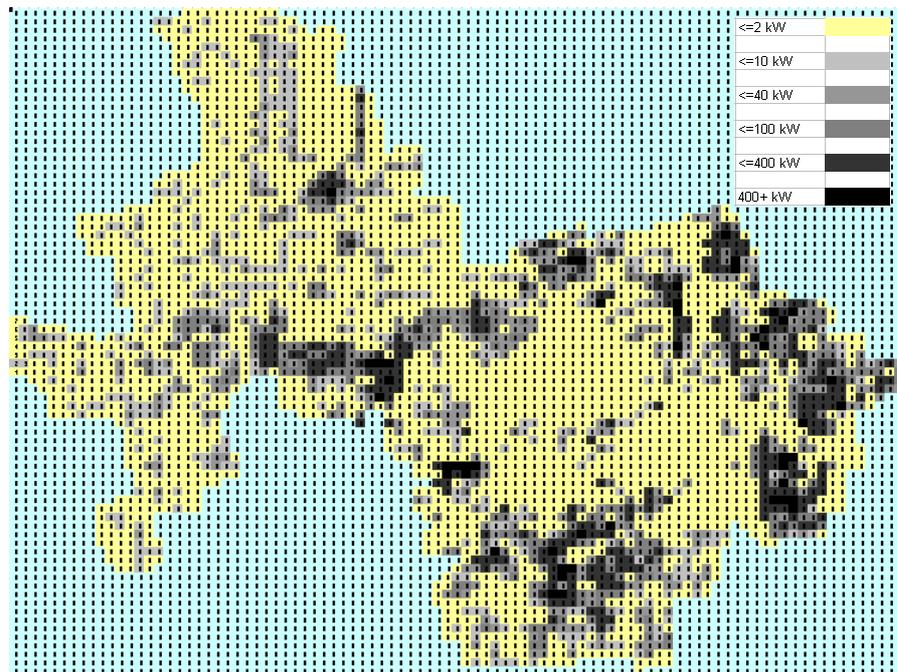


Figure 1.3 Load growth distribution of Madison, WI

The spatial information is addressed by using the small area forecast method: the utility service territory is divided into many, perhaps thousands of small areas, and a forecast is done for each.

Load forecast in utilities can be classified into roughly two categories according to different forecasting needs, short term forecasting and long term forecasting [23]: Short-term forecasting is motivated by a need to reach a decision, to commit to a particular installation or type of construction, and to do so at the lead time. Therefore, the specific aim of the short-term plan is a reliable “alarm” about when present facilities will become insufficient. Risk comparisons (e.g., blackout potential) can be helpful, but they should be used cautiously and tested if possible. There are dangers in comparing risks of diverse character, especially when the intent of the comparison is seen as reducing a risk. One difficulty in using risk comparisons is that it is not always easy to find risks that are sufficiently similar to make a comparison meaningful, because very often one is to compare two alternatives that have two different costs and two different risk levels. The long-term plan evaluates how well the short-term planning commitments fit into long-term needs. No commitment needs to be made to the elements in a long-term plan, and capacity and location are more important than timing in long-term forecast. In other words, it is more important to know *what* will eventually be needed than to know exactly *when* it will be needed.

1.2 Decision Making Process

Based on his/her domain knowledge (e.g., social, economic, or political background), one planner is usually assigned to a certain region in service territory. In every planning period, this planner submits an expansion plan that includes what equipment to maintain, purchase, and install based on his/her experience, e.g., what development activities he/she has seen and heard of in the past, the historical load data and outages that have occurred in this region, the construction-limit land information, etc. After all the planners submit their proposals, the director or higher level managers in charge of planning will make the final decision considering the trade off between benefit and cost, and then allocate a budget to each region accordingly. A comprehensive planning process can be found in Chapter 26 of [23].

Opposite to the practice of budget underestimation for governmental contract competition, most planners tend to overestimate the load to avoid outages in their own regions. The bottom-up aggregation of plans submitted is almost always beyond the total budget. To fulfill the overall budget, which is a hard constraint during the planning stage, overestimation of the expansion plan creates difficulties for the management team to make budget allocation decisions and can potentially be a waste of resources on unnecessary projects. Many utility companies would like to use a computer based program to help examine each plan and execute top-down allocation. When submitted to a state regulator from the utility company, a final plan generated from a standard computer program is usually more convincing and easier to defend than the one purely from planners.

A rather simple technique that can reach absurd conclusions if incorrectly applied is to equate the benefits of a service to the cost of supplying the service by the least expensive alternative method. Thus the benefits from hydroelectricity generation could be estimated as the cost of generating that electricity by the least cost alternative method using solar, wind, geothermal, coal-fired, natural gas or nuclear energy sources. Clearly, this approach to benefit estimation is only valid if, were the project not adopted, the service in question would in fact be demanded at, and supplied by, the least-cost alternative method. The pitfalls associated with this method of benefit and cost estimation can be avoided if one clearly identifies reasonable with and without project scenarios.

The challenge in creating such model-building environments is to make them sufficiently useful and attractive that multiple stakeholders like to use them. They should be understandable. They will have to be relatively easy and transparent, and even fun, to build. They must be capable of simulating and producing different levels of detail with regard to natural, engineering, economic and ecological processes that take place at different spatial and temporal scales. And they must require no programming and debugging by the users. One approach is to develop interactive modeling 'shells' specifically suited to modeling environmental problems. Modeling shells are data-driven programs that become models once sufficient data have been entered into them.

1.3 Literature Review

As load forecasting is highly related to the quality of system planning, attention has been paid to the impact of load forecasting on system design [24] and economics [20]. A system wide two-stage distribution planning algorithm is reported in [19]. Optimization software [4] and techniques [10] are applied to load forecasting as well as planning. Load forecasting is usually tied to reliability analysis [3, 35, 36] and distribution transformer load management (TLM) [5].

There are dozens of different distribution load forecasting methods that have been used and documented during the last 50 years. Some of them fall into the category of short-term forecasting [12, 16], which is beyond the scope of this thesis. Some of them are long-term forecasting methods, the majority of which are spatial load forecasting methods [25, 26]. In [29], the authors summarized and compared 14 different distribution forecasting methods which appeared during the 1960s to early 1980s. Some spatial load forecasting methods can be used for transmission planning as well [11, 34]. As the development of computer technologies and applications ramped up in the 1980s, many computer based load forecasting methods were being developed [17, 21, 33]. Data issues and database development were paid attention to during the same period [22], followed by the discussion of information integration issues in distribution planning [37].

Many methods have been developed to overcome some specific difficulties of spatial load forecasting: fuzzy logic has been applied to forecasting with the consideration of city re-

development [7, 8, 9]; a knowledge-based expert system has been designed for fast developing utility's long-term load forecasting [15]; an extended logistic model was used for high growth load forecasting as well [1]; a method to take care of rural area load forecast was reported in [27]; the load transfer issue has been investigated in [32]; forecasting under uncertainties has been discussed in [31]; neural networks have been applied to forecasting wind speed for long-term load forecasting [2]; the data mining approach has been used for spatial modeling [14]; a load survey system was used to determine customer load characteristics [6]; some fast algorithms have been developed to reduce computer running time [28, 30].

Despite various methods, algorithms, computer codes/programs in use, all fall into three basic types of methods: trending [28], simulation [18] and the hybrid method:

- 1) Trending methods look for some function to fit the past load growth patterns and estimate the future load based on the function. The most common trending method is to use multivariate regression to fit a polynomial function to load history data. This approach has a number of failings when applied to spatial load forecasting, while dozens of improved methods have been reported for load forecasting. The advantages of the trending method include ease of use, simplicity, and a short-range response to recent trends of load growth patterns. However, it often fails to have a useful estimate of the long-range load.
- 2) Simulation methods attempt to model the load growth process to reproduce the load history, as well as to identify the temporal, spatial, and magnitude information of the

future load growth. Simulation methods usually simulate an urban development process based on land-use change information from government, customer rate class from utilities, and load curve model of consumption patterns. Depending on the quality of data, this approach has a fair to very good short-range accuracy. Depending on the specific algorithm, this approach has a good to excellent long-range usefulness for planning. The drawback of the simulation method is the expensive development and training cost.

- 3) Hybrid trending-simulation methods combine features of trending and simulation. An ideal hybrid method should well respond well to recent trends of load history in the short-range, and keep the long-range accuracy as simulation methods have. Meanwhile, the ideal hybrid method should be easy to use, and not require much interaction and skills from the user. That ideal may be unattainable, but it is certainly worth pursuing.

Systems approaches, including the use of extensions of the aforementioned methods can provide an organized framework for resources management and for estimating the important geomorphic, ecological, social and economic impacts and trends over relevant scales of space and time. Within a systems framework, multiple purposes can be investigated, tradeoffs among competing objectives may be identified and evaluated, potential adverse impacts can be assessed, and the various costs and benefits, however measured, of a project may be estimated and examined. This can all be done within a context or process that incorporates the concerns and desires of all those with an interest or stake in the outcome.

Quantitative models can help inform interested stakeholders and those individuals or agencies responsible for recommending or making decisions or policy. The merit and advantages, as well as the limitations, of various quantitative methods for analysing various planning or management issues are generally recognized throughout the community. The assumptions and uncertainties associated with any model-generated impact predictions should be understood and considered by those using these model predictions.

This thesis has been devoted to the investigation of a hybrid method using Gompertz curve (known as S-curve) fitting. S-curve has been used to model diffusion of innovations since 1950s. It has been adopted for transmission & distribution planning since late 1970s, as summarized in [26]. S-curve is typical of a small area, distribution-level load growth, which has three distinct phases: a dormant period (no load or growth in the small area before development), a growth ramp (rapid growth in the small area under construction) and a saturated period (slow growth in the small area being fully developed). There are three parameters that control the shape of an S-curve: horizon year load (HYL, saturated load), time to the start of ramping, and slope of the load growth. As a spatial load forecasting problem has been formulated as an S-curve fitting, the major target is to determine these three parameters, which is a challenging process directly related with the accuracy and usefulness of the forecast. The kernel of the hierarchical trending method includes two hierarchical procedures: bottom-up aggregation and top-down allocation, which are shown in Figure 1.4, the details of which will be presented in Chapter 2.

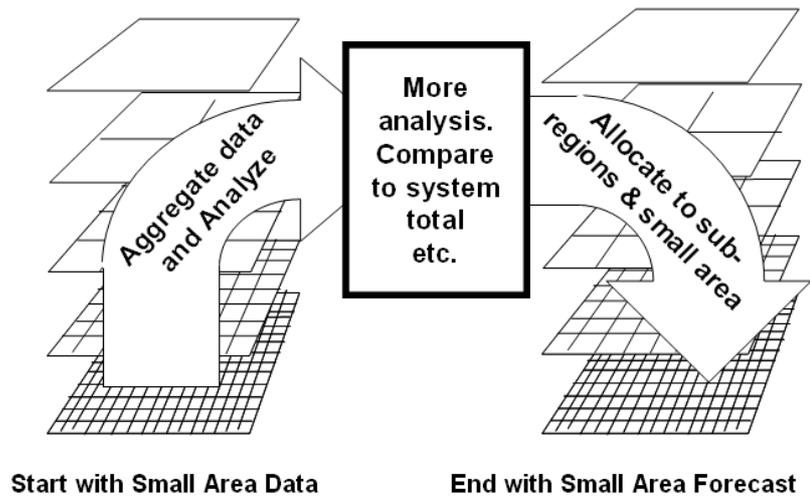


Figure 1.4 Bottom-up aggregation & top-down allocation

A comparison of a computational and human planner's short-term load forecasts was reported in [13]. However, there are rarely discussions of combining human and computer's intelligence. As an enhancement to the proposed hybrid trending method, this thesis also investigates the framework of integrating the two intelligences together to achieve a good solution, which will be discussed in Chapter 3.

1.4 Organization of the Thesis

This thesis includes the whole algorithm development and the Graphical User Interface (GUI) implementation process of a spatial load forecasting tool. Chapter 1 briefly introduces the concept and significance of spatial load forecasting, the decision making process in utilities, and the existing means used to solve spatial load forecasting problems. Chapter 2 presents the detailed development of an automatic hierarchical trending algorithm, where each functional module is described. Chapter 3 introduces an enhancement by embedding human-machine co-construct intelligence (HMCCI) into the tool. Chapter 4 shows the implementation of GUI with load forecasting results. The thesis is summarized in Chapter 5, where future work is discussed as well.

2 Hierarchical Trending Method

2.1 Overall Program Structure

In this chapter, a hierarchical trending method using S-curve has been investigated and developed. Figure 2.1 shows an automated computer program structure.

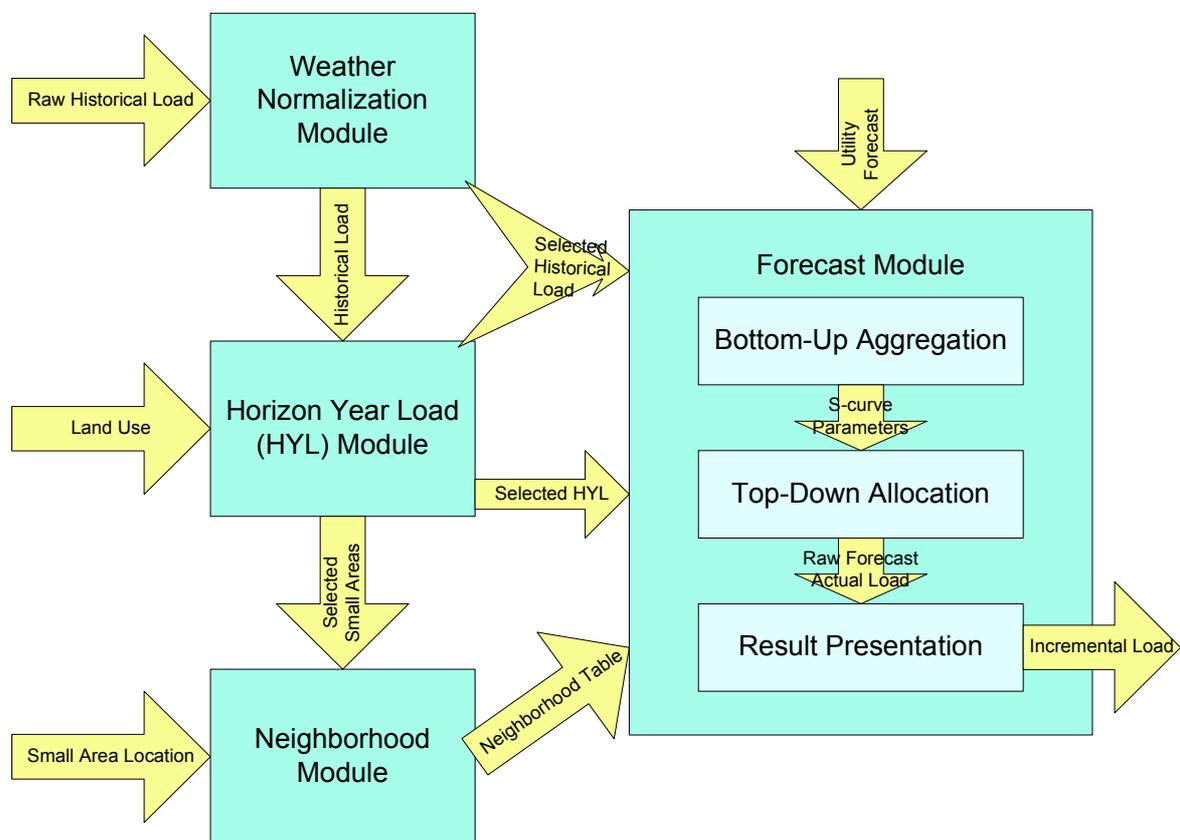


Figure 2.1 An automated computer program structure of hierarchical trending method

As shown in Figure 2.1, the arrows represent data flow, while the rectangular blocks represent functional modules:

- 1) Weather normalization module: this module applies a utility's weather normalization method to the raw historical load data to generate the adjusted historical load. Different utility companies may have different weather normalization methods, which mainly fall into various regressions.
- 2) Horizon year load (HYL) module: this module takes land use data and adjusted historical load data to generate load densities for each land use type as well as the HYL for each small area. It then selects the small areas with higher HYL than the current load as the areas of interest to forecast the load growth on.
- 3) Neighborhood module: this module builds the neighborhood table according to the total number and the location, sometimes together with load information as well, of the selected areas of interest.
- 4) Forecast module: this module includes three sub-modules. The bottom-up module iteratively fits S-curves and aggregates historical load and HYL for each small area in each level based on the neighborhood table. The top-down module takes the S-curve parameters from the bottom-up module as references, and allocates the utility's system forecast from the top level to the bottom level. Finally, the result representation module fine tunes the raw forecasting results and displays the forecast load in both data sheets and map format.

The basic load forecast tool is made up of the four modules mentioned above together with some display and user interface functions. Each module will be discussed in the following sections, and the enhancement of the tool embedded with the HMCCI framework will be discussed in the next chapter.

2.2 Data Outlook

Compared with simulation data, real-world data provided by utilities has a lot more uncertainties, noise and errors, which create difficulties for the forecasting. Before starting the forecast algorithm development, several typical data sets obtainable from utilities are introduced in this section.

2.2.1 Historical Load

Figure 2.2 shows the data format of transformer based historical load information that most utilities can provide, which contains the following two sets:

- 1) Geographical information: In the geographical information system (GIS), the surface of the earth has been cut into many small areas according to a certain scheme (e.g., 1500ft by 1500ft). Each small area has a unique shape number, which is shown in column A of Figure 2.2. The location of a transformer in the small area can be represented by latitude and longitude, or another coordination scheme specified by a particular GIS. As shown in Figure 2.2, Columns B and C represent the longitude and latitude of the corresponding transformer, while Columns D and E represent its coordinate in feet.
- 2) Load history: Load history information and details of the transformer (e.g., feeder number, voltage level, etc.) are listed after geographical information, from which the load history on a small area basis will be extracted.

1	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	
1	SHAPE_NUMBER_1600	LATITUDE	LONGITUDE	SDO_Y1	SDO_Y1	G3E_FID		CURRENT_FEEDE	Y2000SYNERGEESECTION	Y2000FEEDEID	Y2000PARENTFEEDE	Y2000TR	Y2000SUB	Y2000PHASEKV_CO	Y2000PHASEKV_CC_Y2	
2	71951	-89.3884856	43.0787068	2163379.51	393749.14	395996	1-A7	HKV ECA Network	NM		HKV ECA Network	0	ECA TRI	ECA	1615391325	1615391325
3	71951	-89.388399	43.0799879	2163399.22	394216.227	1264622	1-E10	BLT 415	Seg: 11930 - 11931	BLT 415	0	BLT RAD TR6	BLT4	22.88227324	14.68579045	
4	71951	-89.3871153	43.0795488	2163743.35	394058.691	1264619	1-E12	BLT 415	Seg: 11937 - 11938	BLT 415	0	BLT RAD TR6	BLT4	21.39106892	13.7004231	
5	71951	-89.3875737	43.0798444	2163629.87	394185.615	1264619	1-E13	BLT 415	Seg: 11937 - 11939	BLT 415	0	BLT RAD TR6	BLT4	9.204330136	5.899223013	
6	71951	-89.3877303	43.0802452	2163577.16	394311.299	127474	1-E7	BLT 415		0	0	0	0	0	0	
7	71951	-89.3862247	43.0802951	2163979.26	394333.526	1264610	1-E8	HKV BLT 1006	NM		HKV BLT Network	0	BLTH TR1 TR2	BLTH	200.9193	200.9193
8	71951	-89.3860415	43.0794345	2164297.69	394021.107	131935	1-C3	BLT 415	Seg: 11933 - 11934	BLT 415	0	BLT RAD TR6	BLT4	0	11.5679482	
9	71951	-89.3862418	43.0785816	2163979.29	393707.915	129476	1-C5	BLT 415	Seg: 11915 - 12093	BLT 415	0	BLT RAD TR6	BLT4	0	10.84279788	
10	71951	-89.3874508	43.0788892	2163655.47	393817.653	0	1-D4		0	Seg: 11929 - 11930	BLT 415	0	BLT RAD TR6	BLT4	0	45.21677415
11	72181	-89.3817531	43.0818753	2165169.65	394917.185	131933	10-A10	BLD 1318	Seg: 17523 - 17740	BLT 413	0	BLT RAD TR5	BLT4	0	0	
12	72181	-89.3792019	43.08042	2165844.89	395769.474	131955	10-B2	BLD 1318	Seg: 14026 - 14027	BLT 413	0	BLT RAD TR5	BLT4	0	0	
13	72181	-89.3788102	43.0838265	2165950.54	395634.128	131924	10-B3	BLD 1318	Seg: 14020 - 14026	BLT 413	0	BLT RAD TR5	BLT4	0	0	
14	72181	-89.3795425	43.0833199	2165756.3	395446.228	4256756	10-B5	BLD 1318		0	0	0	0	0	0	
15	72181	-89.3798362	43.0831093	2165678.4	395370.347	131922	10-B6	BLD 1318	Seg: 14024 - 14025	BLT 413	0	BLT RAD TR5	BLT4	0	0	
16	72181	-89.3804403	43.0830792	2165517.09	395358.531	2306509	10-B9	BLD 1318	Seg: 14024 - 17524	BLT 413	0	BLT RAD TR5	BLT4	21.9499473	23.83793235	
17	71722	-89.3915969	43.0785472	2162548.7	393694.916	605706	101-A24	HKV ECA Network	NM		HKV ECA Network	0	ECA TRI	ECA	12.38580534	12.38580534
18	71722	-89.3912114	43.0783019	2162692.34	393696.242	605712	101-A25	HKV ECA Network	NM		HKV ECA Network	0	ECA TRI	ECA	12.00986912	12.00986912
19	71722	-89.3916671	43.0776624	2162532.27	393382.291	605825	101-A26	HKV ECA Network	NM		HKV ECA Network	0	ECA TRI	ECA	65.95553547	65.95553547
20	71722	-89.3928811	43.0778756	2162207.43	393429.635	440953	101-A27	HKV ECA Network	NM		HKV ECA Network	0	ECA TRI	ECA	18.66539564	18.66539564
21	71722	-89.3903048	43.0783778	2162894.35	393625.675	608164	101-B4	HKV ECA Network	NM		HKV ECA Network	0	ECA TRI	ECA	30.09490958	30.09490958
22	71722	-89.3902006	43.0781494	2162920.12	393907.096	609692	101-B5	HKV ECA Network	NM		HKV ECA Network	0	ECA TRI	ECA	108.7159796	108.7159796
23	71722	-89.3901007	43.0779793	2162942.94	393341.786	144377	101-D1	BLT 415	Seg: 11926 - 10609	BLT 415	0	BLT RAD TR6	BLT4	0	20.00479731	
24	71721	-89.3917688	43.076478	2162508.25	392930.448	609508	101-D10	HKV ECA Network	NM		HKV ECA Network	0	ECA TRI	ECA	69.63331903	69.63331903
25	71722	-89.3902593	43.0768222	2162911.71	393058.842	1269188	101-E12	BLT 415	Seg: 11928 - 11927	BLT 415	0	BLT RAD TR6	BLT4	38.35994587	24.68579702	
26	71951	-89.38935	43.0770034	2163153.11	393126.637	129557	101-E3	BLT 415	Seg: 11922 - 11925	BLT 415	0	BLT RAD TR6	BLT4	10.23274691	6.559409662	
27	71721	-89.3904277	43.0756385	2162868.81	392627.112	613903	101-E6	BLT 415	Seg: 11923 - 11924	BLT 415	0	BLT RAD TR6	BLT4	0	32.2378216	
28	71951	-89.3894547	43.077611	2163123.78	393311.442	129545	101-F1	BLT 415	Seg: 11926 - 11926	BLT 415	0	BLT RAD TR6	BLT4	0	19.97184048	
29	71951	-89.3882072	43.0775548	2163456.95	393329.84	395995	101-F9	4KV BLT Network	NM		4KV BLT Network	0	BLT NET TR3 TF BLTNET	BLT4	46.19591454	46.19591454
30	71950	-89.3898379	43.0758088	2163025.47	392753.032	129487	102-A1	BLT 415	Seg: 11922 - 11923	BLT 415	0	BLT RAD TR6	BLT4	0.359946871	0.230697827	
31	71950	-89.3889783	43.0765885	2163253.53	392974.881	294121	102-A9	BLT 415	Seg: 11921 - 11922	BLT 415	0	BLT RAD TR6	BLT4	0	0	
32	71721	-89.390447	43.0751832	2162864.87	392461.139	605724	102-A9	HKV ECA Network	NM		HKV ECA Network	0	ECA TRI	ECA	16.40405693	16.40405693
33	71951	-89.3879127	43.0766997	2163537.91	393018.789	395992	102-B5	4KV BLT Network	NM		4KV BLT Network	0	BLT NET TR3 TF BLTNET	BLT4	70.53974721	70.53974721
34	71950	-89.3872974	43.0764369	2163703.03	392924.21	127354	102-C5	4KV BLT Network	NM		4KV BLT Network	0	BLT NET TR3 TF BLTNET	BLT4	236.2467254	236.2467254
35	71950	-89.3876001	43.0763162	2163524.71	392967.005	127354	102-C6	4KV BLT Network	NM		4KV BLT Network	0	BLT NET TR3 TF BLTNET	BLT4	22.92910661	22.92910661

Figure 2.2 Typical data format of transformer based historical load

2.2.2 Weather Normalization

Historical data is measured under various conditions including different temperatures, economic situations and so forth, which influence the load consumption. A theoretical load consumption curve should be homogeneously increasing, while in reality, this curve turns out to zigzag with time. In order to compare an apple to an apple, a normalization procedure has to be conducted before the load forecasting process begins. Up to now, there is no identified method to do weather normalization among the power industries, and different utility companies may use different means, from simple to complicated, according to their local situation. In this thesis, a multivariate regression approach will be discussed. The required data format is shown in Figure 2.3. Other than the actual peak load for each year, three other columns of data are required for weather normalization:

- 1) Maximum temperature records the highest temperature that appeared during the year of interest.
- 2) Cooling Degree Days (CDD) describes how much a period's weather should result in a building's cooling requirements. The hotter the day is, the more the CDD is. If the amount of CDD is double, then this should result in roughly double the cooling requirements for a building. CDD is calculated individually for each day. CDD over a month or billing cycle are merely a summation of CDD of the individual days.
- 3) Employment shows the number of employees or job positions in the service territory. Sometimes it can be replaced by number of customers if the utility company has customer count data in the service territory.

Year	Actual Peak Load	Maximum Temperature	Cooling Degree Days	Employment (or Customers)
1988	518.01	101.49	23.10	108.62
1989	512.86	94.63	18.92	110.06
1990	518.01	91.84	16.95	111.59
1991	523.22	92.76	17.99	112.86
1992	482.99	89.12	9.03	114.22
1993	539.15	90.02	16.95	115.89
1994	550.04	93.69	15.96	117.73
1995	632.70	101.49	25.03	119.33
1996	584.06	91.84	17.99	120.75
1997	589.93	90.02	12.94	122.06
1998	607.89	87.36	11.94	123.26
1999	692.29	79.04	20.09	125.57
2000	665.14	90.92	15.96	126.15
2001	713.37	91.84	20.91	127.44
2002	692.29	90.92	18.92	129.53
2003	699.24	93.69	16.95	132.25
2004	632.70	87.36	15.03	133.89
2005	699.24	92.76	15.96	135.10
2006	742.48	94.63	20.91	136.66
2007	685.40	90.92	12.94	137.56

Figure 2.3 Typical data format for weather normalization

2.2.3 Horizon Year Load

A land use based approach to calculate horizon year load (HYL) will be proposed in section 2.4. This approach requires a land use code and current and future land use information:

A land use code describes the purpose of land use or customer types within the small area, e.g., commercial, industrial, low-density residential, high-density residential, water body, etc.

All types are coded using numbers as shown in Figure 2.4.

C	D
10	Water Body
20	Industrial / Business
39	Extractive
40	Transportation
50	Commercial (Retail and Services)
70	Institutional / Government
80	Parks / Outdoor Recreation
90	Agriculture / Vacant
97	Natural Area
99	Woodland
100	Rural Residential
110	Low Density Residential
120	Medium Density Residential
130	High Density Residential
200	Mixed Commercial / Residential
210	Planned Neighborhood
400	Communication / Utilities
999	Under Construction

Figure 2.4 Land use code

Current and future land use present the exact amount of land used for each land use code on a small area basis, e.g., in a small area, 21.3% of the area is used for commercial purposes, 10.8% of the area is a water body (lake or river), and the other part is vacant. Both current

and future land use information are required to calculate HYL. The data format of current land use is shown in Figure 2.5, while future land use has exactly the same format as shown in Figure 2.6.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	CURRENT LAND USE TYPE													
2	1	2	3	4	5	6	7	8	9	10	11	12	13	
3	SHAPE NUMBER	Industrial / Business	Extractive	Commercial (Retail & Services)	Institutional / Government	Rural Residential	Low Density Residential	Medium Density Residential	High Density Residential	Communication / Utilities	Transportation	Agriculture / Vacant	Parks / Outdoor Recreation	Natural / Woodland / Water / Other
4	57530	0	0	0	0	0	0	0	0	0	0	0	0	0.400239
5	57531	0	0	0	0	0	0.04697	0	0	0	0.031715	0.07542	0	0.779836
6	57532	0	0	0	0	0	0	0	0	0	0.010283	0.3479	0	0.562368
7	57533	0	0	0	0	0	0	0	0	0	0	0.219471	0	0.680182
8	57534	0	0	0	0	0	0	0	0	0	0	0.639908	0	0.16667
9	57535	0	0	0	0	0	0	0	0	0	0	0.429012	0	0
10	57536	0	0	0	0	0	0	0	0	0	0	0.021801	0	0
11	57759	0	0	0	0	0	0	0	0	0	0	0.00798	0	0.730754
12	57760	0	0	0	0	0	0.01473	0	0	0	0.00332	0.52495	0	0.456966
13	57761	0	0	0	0	0	0.003938	0	0	0	0.040432	0.805305	0	0.150192
14	57762	0	0	0	0	0	0	0	0	0	0	0.408337	0	0.591663
15	57763	0	0	0	0	0	0.047528	0	0	0	0	0.511942	0	0.440531
16	57764	0	0	0	0	0	0.041268	0	0	0	0.033461	0.924399	0	0
17	57765	0	0	0	0	0	0	0	0	0	0.004239	0.041426	0	0
18	57987	0	0	0	0	0	0	0	0	0	0	0.11824	0	0.000128
19	57988	0	0	0	0	0	0	0	0	0	0	0.551111	0	0.430343
20	57989	0	0	0	0	0	0.01056	0	0	0	0.052258	0.799434	0	0.137489
21	57990	0	0	0	0	0	0.034624	0	0	0	0.077841	0.594844	0	0.292621
22	57991	0	0	0	0	0	0	0	0	0	0	0.714376	0	0.285624
23	57992	0	0	0	0	0	0.194692	0	0	0	0.029312	0.274703	0	0.501282
24	57993	0	0	0	0	0	0.076351	0	0	0	0.06076	0.696614	0	0.092408
25	58215	0	0	0	0	0	0	0	0	0	0.00056	0.003078	0	0.002298
26	58216	0	0	0	0	0	0	0	0	0	0	0.341557	0	0.202446

Figure 2.5 Current land use data

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	FUTURE LAND USE TYPE													
2	1	2	3	4	5	6	7	8	9	10	11	12	13	
3	SHAPE NUMBER	Industrial / Business	Extractive	Commercial (Retail & Services)	Institutional / Government	Rural Residential	Low Density Residential	Medium Density Residential	High Density Residential	Communication / Utilities	Transportation	Agriculture / Vacant	Parks / Outdoor Recreation	Natural / Woodland / Water / Other
10	57536	0	0	0	0	0	0	0	0	0	0	0.021801	0	0
11	57759	0	0	0	0	0	0	0	0	0	0	0.007981	0	0.730754
12	57760	0	0	0	0	0.01473	0	0	0	0	0.00332	0.52495	0	0.456966
13	57761	0	0	0	0	0.003938	0	0	0	0	0.040523	0.805305	0	0.150192
14	57762	0	0	0	0	0	0	0	0	0	0	0.408337	0	0.591663
15	57763	0	0	0	0	0.047528	0	0	0	0	0	0.511942	0	0.440531
16	57764	0	0	0	0	0.041268	0	0	0	0	0.033462	0.924399	0	0
17	57765	0	0	0	0	0	0	0	0	0	0.004239	0.041426	0	0
18	57987	0	0	0	0	0	0	0	0	0	0	0.11824	0	0.000128
19	57988	0	0	0	0	0	0	0	0	0	0	0.551111	0	0.430343
20	57989	0	0	0	0	0.010561	0	0	0	0	0.052237	0.799436	0	0.137489
21	57990	0	0	0	0	0.034622	0	0	0	0	0.077938	0.594843	0	0.292602
22	57991	0	0	0	0	0	0	0	0	0	0	0.714376	0	0.285624
23	57992	0	0	0	0	0.194692	0	0	0	0	0.029357	0.274707	0	0.501278
24	57993	0	0	0	0	0.076351	0	0	0	0	0.06077	0.696613	0	0.092408
25	58215	0	0	0	0	0	0	0	0	0	0.000566	0.003078	0	0.002298
26	58216	0	0	0	0	0	0	0	0	0	0	0.341557	0	0.202446
27	58217	0	0	0	0	0	0	0	0	0	0	0.84529	0	0.154697
28	58218	0	0	0	0	0.021588	0	0	0	0	0	0.178347	0	0.800331
29	58219	0	0	0	0	0.022824	0	0	0	0	0.073174	0.233493	0	0.670388
30	58220	0	0	0	0	0	0	0	0	0	0.019828	0.437854	0	0.542253

Figure 2.6 Future land use data

2.3 Data Preprocessing

2.3.1 Small Area Based Historical Load

The first step of data preprocessing is to convert transformer based load data to small area based load data and assign coordinates to each small area. The extraction procedure is the following:

- 1) Rank the shape numbers of all the small areas from low to high;
- 2) Define the reference point for each small area, e.g., the upper-left corner;
- 3) Add the load of transformers with the same shape number together, the result of which is the historical load for this small area.

As shown in Figure 2.7, column B and C show the coordinates of the reference point in a small area, while columns D to J are small area based annual peak load data from 2001 to 2007, the unit of which is kW.

1	A	B	C	D	E	F	G	H	I	J
2	SHAPE NUMBER	COORDINATE X	COORDINATE Y	2001	2002	2003	2004	2005	2006	2007
3	57530	2068500	402000	0	0	0	0	0	0	0
4	57531	2068500	403500	5.47191	4.46938	5.53405	5.2614	6.03313	3.32276	3.35085
5	57532	2068500	405000	0	0	0	0	0	0	0
6	57533	2068500	406500	0	0	0	0	0	0	0
7	57534	2068500	408000	0	0	0	0	0	0	0
8	57535	2068500	409500	0	0	0	0	0	0	0
9	57536	2068500	411000	0	0	0	0	0	0	0
10	57759	2070000	402000	0	0	0	0	0	0	0
11	57760	2070000	403500	8.84625	7.16416	10.8739	9.03373	8.46369	4.53365	5.41002
12	57761	2070000	405000	0	0	0	0	0	0	0
13	57762	2070000	406500	0	0	0	0	0	0	0
14	57763	2070000	408000	9.94063	7.75569	10.971	9.62936	16.5619	9.02389	9.60405
15	57764	2070000	409500	4.10393	3.28631	4.36899	3.07742	3.24155	3.84214	2.88834
16	57765	2070000	411000	0	0	0	0	0	0	0
17	57987	2071500	400500	0	0	0	0	0	0	0
18	57988	2071500	402000	0	0	0	0	0	0	0
19	57989	2071500	403500	1.82397	1.31452	2.03886	1.88616	2.2424	1.04988	1.11901
20	57990	2071500	405000	12.3118	10.2533	15.7284	15.1886	17.3415	9.53469	9.78738
21	57991	2071500	406500	0	0	0	0	0	0	0
22	57992	2071500	408000	14.683	11.4364	11.6506	10.6221	12.3635	7.1432	5.95495
23	57993	2071500	409500	1.09438	22.2155	28.2528	24.9172	32.0002	10.2585	13.5677

Figure 2.7 Small area based historical load data

2.3.2 Service Territory Map

According to the coordinates of small areas, a service territory map can be drawn as the light yellow area in Figure 2.8. In the beginning stage, the service territory map can be used to verify the data quality. Sometimes due to errors from a data source, some regions may not appear in the map correctly. By comparing the generated map with the real map, one can tell approximately whether the service territory has been correctly represented.



Figure 2.8 Service territory map

2.3.3 Weather Normalization

In this thesis, a system-wide normalization is used to generate coefficients to normalize the small area historical load. As shown in Figure 2.3, four categories of data will be used in a system-wide normalization: actual peak load (y), maximum temperature (x_1), CDD (x_2),

and employment (x_3), the natural log of which are denoted as y , x_1 , x_2 , and x_3 respectively.

The least square method is used to calculate the coefficient (k_1 , k_2 , and k_3) of a line to best fit the given data, the formula of which is

$$y = k_1x_1 + k_2x_2 + k_3x_3 \quad (2.1)$$

The weather normalized system load (y') is shown as the blue curve in Figure 2.9, while the purple curve represents the actual load.

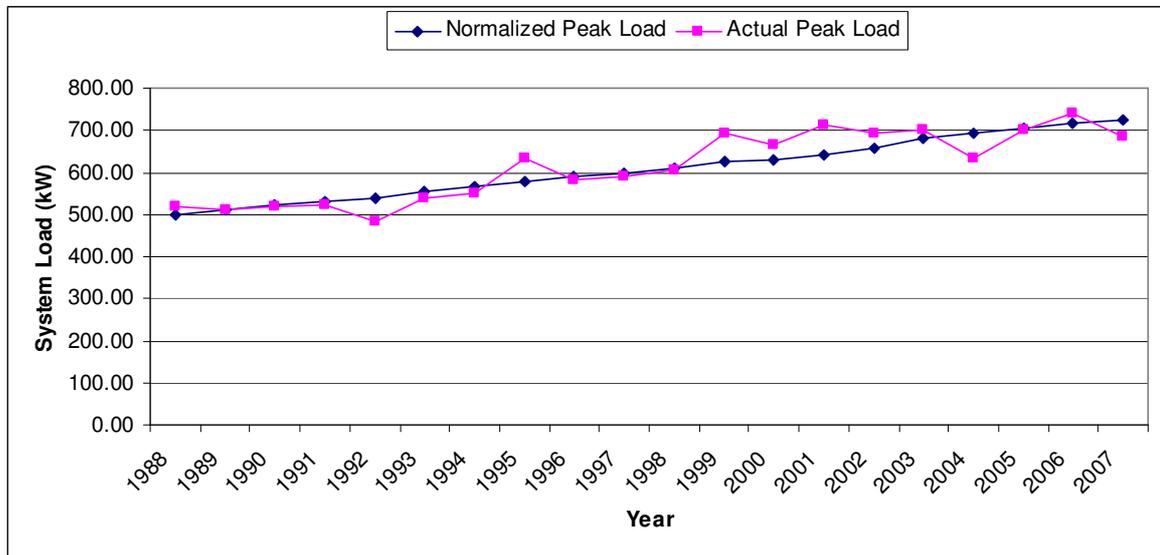


Figure 2.9 Weather normalized load vs. actual load

The coefficient c_i is used to normalize small area load of a particular year i , and can be calculated by:

$$c_i = \frac{y'_i}{y_i}, \quad (2.2)$$

where y_i is the actual system load of year i , and y'_i is the normalized system load of the same year.

And the normalized load is equal to the actual historical load data multiplied by the coefficient c_i :

$$L'_{ij} = c_i L_{ij}, \quad (2.3)$$

where L_{ij} is the actual historical load of a small area j in year i , and L'_{ij} is the normalized load.

2.4 Determining Horizon Year Load Using Land Use Information

Horizon year load (HYL) describes the load of a fully saturated small area. In long-range spatial load forecast, an estimate of the farthest forecast year load can be considered as HYL, although the land may be further developed or redeveloped after the forecast range. It is crucial to get a quality HYL for a useful forecast. Figure 2.10 and Figure 2.11 show how different HYLs affect the forecast given the same historical data.

As shown in Figure 2.10, in which HYL increases from 10kW to 80kW, the long range forecast varies almost linearly according to the change of HYL. Historical data takes the role of the saturation period when HYL is 10kW, while it acts as a dormant period when HYL is 80kW. Figure 2.11 shows how a small variation of HYL affects the forecast. When HYL varies from 10kW to 16kW, historical data stays in the saturated stage, while the forecast of the recent years vary less significantly than years further out.

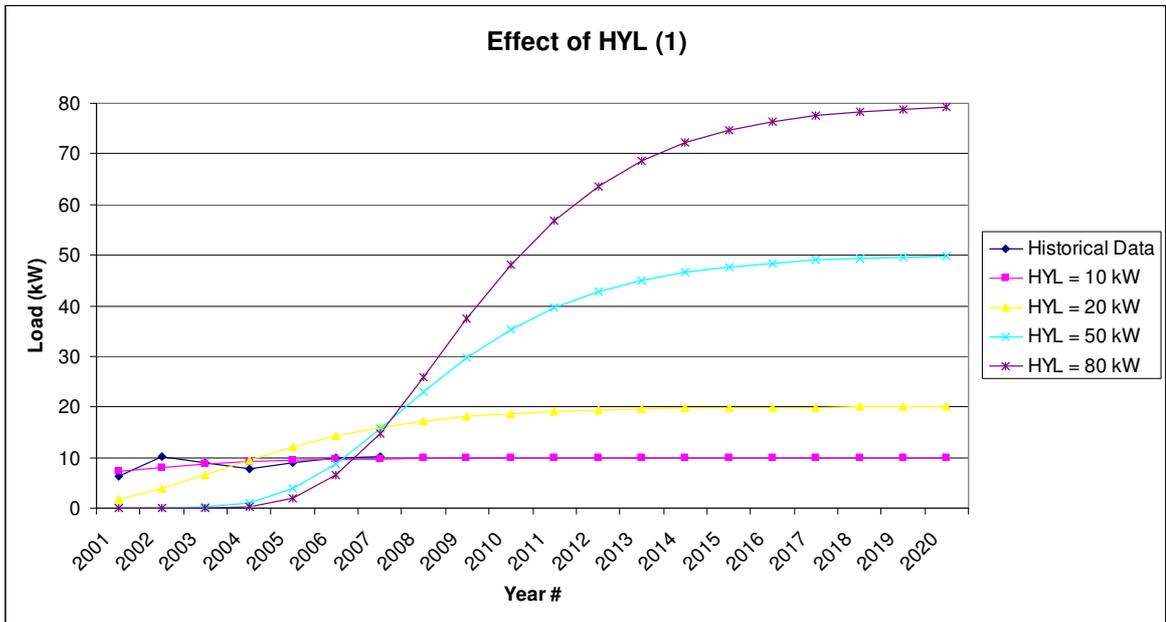


Figure 2.10 Effect of HYL on the overall forecast (1)

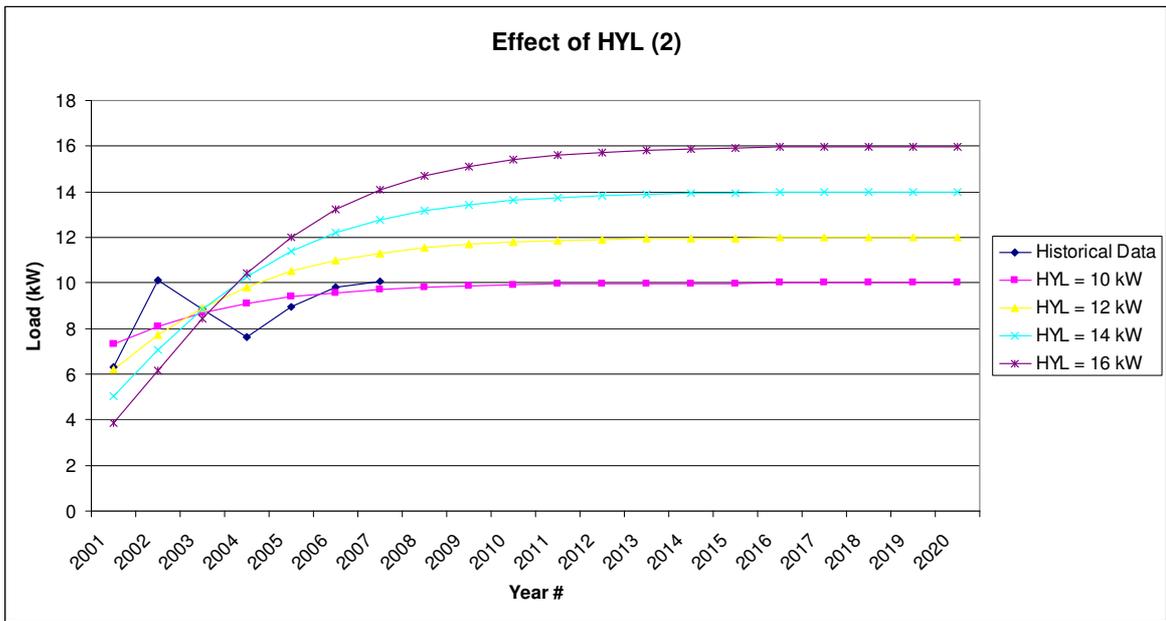


Figure 2.11 Effect of HYL on the overall forecast (2)

HYL of the small areas in the service territory can be calculated based on land use data from the county, customer count data from utilities, or both. In this section, a land use based

method to determine HYL is introduced. An implementation in MS Excel is described as follows:

The land use types are divided into 10 to 15 categories based on physical meaning of data (e.g., group different factories as industrial land use). Original land use data from a given county may includes 20 or even more types, which is inefficient for calculation, while 10 to 15 categories are fewer and more practical considering both the computational complexity and the quality of the results.

The load density variables (d_k , $k = 1, 2, 3\dots$) are set for each category, and are used to calculate the load of the most recent historical year (known as base year b) for each small area by multiplying load density with current land use:

$$\hat{L}_{bj} = \sum_k (d_k \cdot CLU_{kj}), \quad (2.4)$$

where \hat{L}_{bj} is the calculated base year load of small area j , d_k is the load density of category k , and CLU_{kj} is the acreage of current land use category k in small area j .

To solve the optimization formulation using Excel add-in Solver, tune the load density variables to minimize the sum of square errors between the calculated base year load and the weather normalized base year load with the constraint of non-negative load density:

$$\text{Min:} \quad \sum_j (\hat{L}_{bj} - L'_{bj})^2$$

$$\text{s.t.} \quad d_k \geq 0, k = 1, 2, 3, \dots \quad (2.5)$$

Calculate the mismatch E_j between the calculated and weather normalized base year load:

$$E_j = \hat{L}_{bj} - L'_{bj}, \quad (2.6)$$

Multiply the calculated load density with future land use, and then add the mismatch calculated from step 4 to the result from step 5 to get the adjusted load of horizon year, which is HYL.

$$HYL_j = \sum_k (d_k \cdot FLU_{kj}) + E_j, \quad (2.7)$$

where FLU_{kj} is the acreage of future land use category k in small area j .

Notice that in many cases, several small areas with a large mismatch may have a big affect on the load densities. In practice, it is good to treat this small proportion of data as outliers and use the other small areas to calculate the load densities.

2.5 Neighborhood Module

In the hierarchical trending method, small areas are aggregated from bottom to top and the system-wide load is distributed from top to bottom. During this process, small areas within the same neighborhood are aggregated, and the load of the upper level region will be allocated to the small areas within the same neighborhood as well. Neighbor here may not be geographically adjacent, but a scheme exists to group the small areas or sub regions in each level. The neighborhood table (Figure 2.12) is the link of this iterative method. This section introduces a simple but robust approach to build a neighborhood table.

	A	B	C	D	E	F
1	1	1	1	1	1	1
2	2	1	1	1	1	1
3	3	1	1	1	1	1
4	4	1	1	1	1	1
5	5	1	1	1	1	1
6	6	2	1	1	1	1
7	7	2	1	1	1	1
8	8	2	1	1	1	1
9	9	2	1	1	1	1
10	10	2	1	1	1	1
11	11	3	1	1	1	1
12	12	3	1	1	1	1
13	13	3	1	1	1	1
14	14	3	1	1	1	1
15	15	3	1	1	1	1
16	16	4	1	1	1	1
17	17	4	1	1	1	1
18	18	4	1	1	1	1
19	19	4	1	1	1	1
20	20	4	1	1	1	1
21	21	5	2	1	1	1
22	22	5	2	1	1	1
23	23	5	2	1	1	1
24	24	5	2	1	1	1
25	25	5	2	1	1	1
26	26	6	2	1	1	1
27	27	6	2	1	1	1
28	28	6	2	1	1	1
29	29	6	2	1	1	1
30	30	6	2	1	1	1
31	31	7	2	1	1	1
32	32	7	2	1	1	1
33	33	7	2	1	1	1

Figure 2.12 Neighborhood table

The neighborhood table is a table of indices of sub-regions. As shown in Figure 2.12, the columns represent the levels from bottom to top, and each entry represents the sub-region index in the corresponding level. For example, column B represents the 2nd level bottom-up, and cell B7 represents the 2nd region in level #2. In this neighborhood table, small areas #1 through #5 in level #1 form sub-region #1 in level #2, sub-regions #1 through #4 in level #2 form sub-region #1 in level #3, and so forth. In each level, every sub-region should include the same amount (4 or 5) of sub-regions in the lower level, though this is not a hard requirement.

With the above considerations, the procedure of building a neighborhood table is described as follows:

- 1) Determine a load growth threshold as the just noticeable difference for load forecasting. In other words, if the horizon year load is no more than the sum of this threshold and the base year load, planners can consider that there is no growth in this small area through the forecast range.
- 2) Build a binary tree to calculate the potential numbers of sub-regions in each level as shown in Figure 2.13. Each child node is equal to its parent node multiplied by 4 (if it is the left child node) or 5 (if it is the right child node). The binary tree is built using the depth-first search from the right to the left. Eliminate the node if its value is the same as the one generated already. The value of the node in the bottom level should be no more than the number of small areas.

- 3) Determine the number of small areas that will be considered for forecast. This number is the smallest potential sub-region number that includes all small areas whose mismatches are above the threshold.
- 4) Assign indices to the selected small areas. In the binary tree, trace from level #1 to the root to build the neighborhood table using the multiplier (4 or 5).

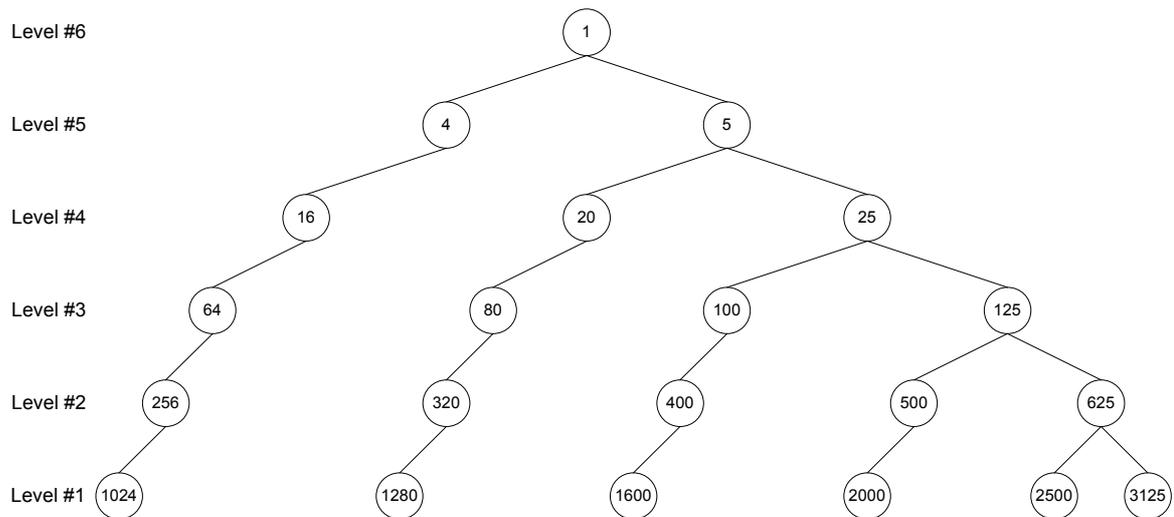


Figure 2.13 Potential amount of sub-regions in each level

2.6 Bottom-Up Aggregation: S-curve Parameter Tuning

Historical load and horizon year load of the small areas within each neighborhood are aggregated from the bottom level to the top level as shown in Figure 2.14. In each level, S-curve is used to fit the corresponding data from each small area (bottom level) or sub-region (second level or higher). During this bottom-up aggregation, the S-curve fitting is a crucial segment, the results of which will be the starting point for each small area or sub-region of top-down allocation. Parameter tuning of S-curve will be discussed in the following sections.

2.6.1 Format of S-curve

The typical format of S-curve (Gompertz function) can be written as:

$$y(t) = ae^{be^{ct}}, \quad (2.8)$$

where a is the upper asymptote, c is the growth rate, and b and c are negative numbers. This typical format can be adopted as:

$$y(t) = ae^{-e^{c(t+\ln(-b))/c}}. \quad (2.9)$$

Let

$$\Delta t = -\frac{1}{c} \ln(-b). \quad (2.10)$$

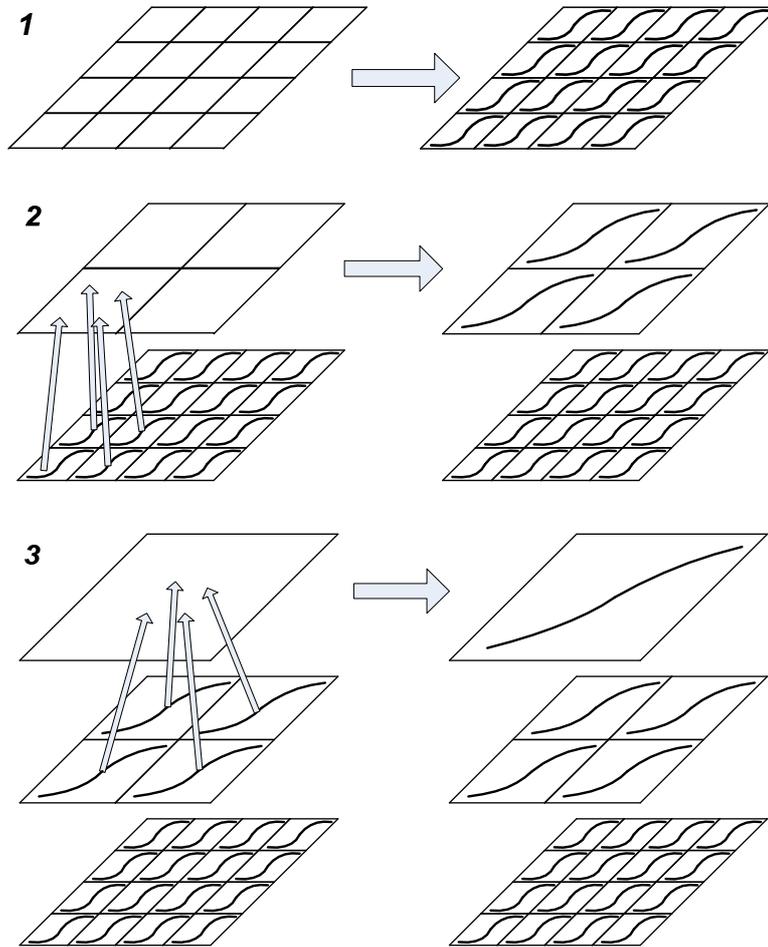


Figure 2.14 Bottom-up aggregation

Formula (2.9) is equivalent to:

$$y(t) = ae^{-e^{c(t-\Delta t)}}, \quad (2.11)$$

where a is the HYL of a small area, c is the load growth rate, and Δt is the ramp up time.

Comparing with the typical format in (2.8), formula (2.11) is fundamentally the same but has the advantage that each of the three parameters has its own physical meaning, which is

important to the users (planners) because the parameter tuning process can be easily visualized. Therefore, formula (2.11) has been used to represent S-curve in this thesis. For example, the ramp up time depends upon both b and c in (2.8), while in (2.11) it is determined by Δt only.

Figure 2.15 shows how the shape of S-curve is affected by tuning HYL parameter a : the S-curve is stretched vertically as the HYL parameter is increasing. Figure 2.16 shows how the shape of S-curve is affected by slope parameter c : the S-curve is stretched horizontally as the absolute value of c is decreasing. As the slope is approaching zero, the S-curve tends to become a straight line. Figure 2.17 shows how the shape of S-curve is affected by ramp up time parameter Δt : the S-curve moving from the left to the right as Δt is increasing.

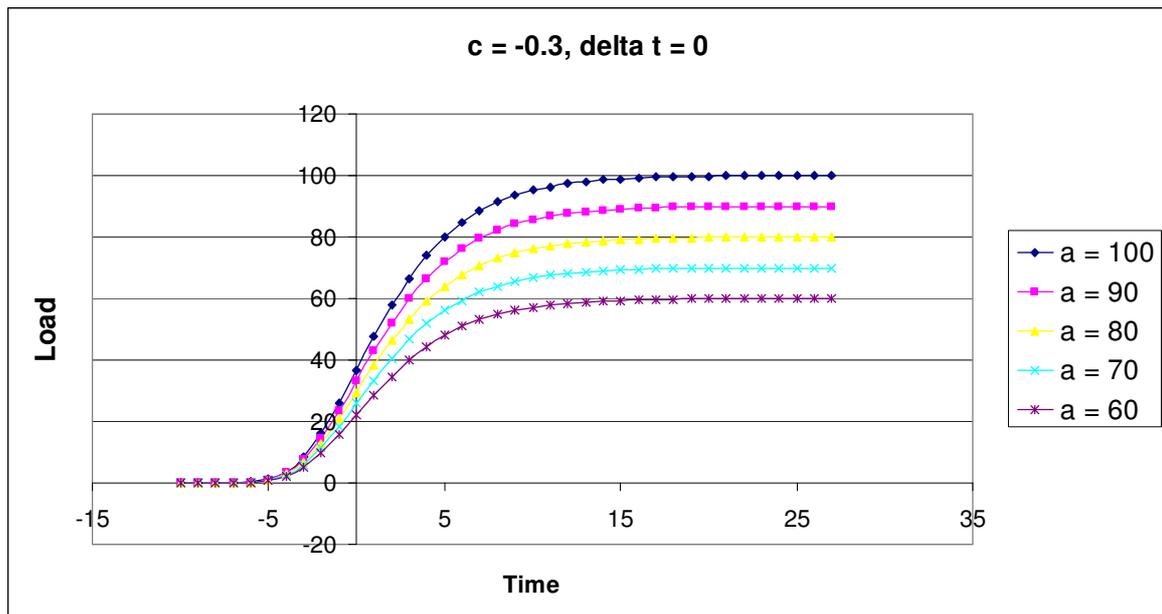


Figure 2.15 S-curve shapes affected by HYL parameter

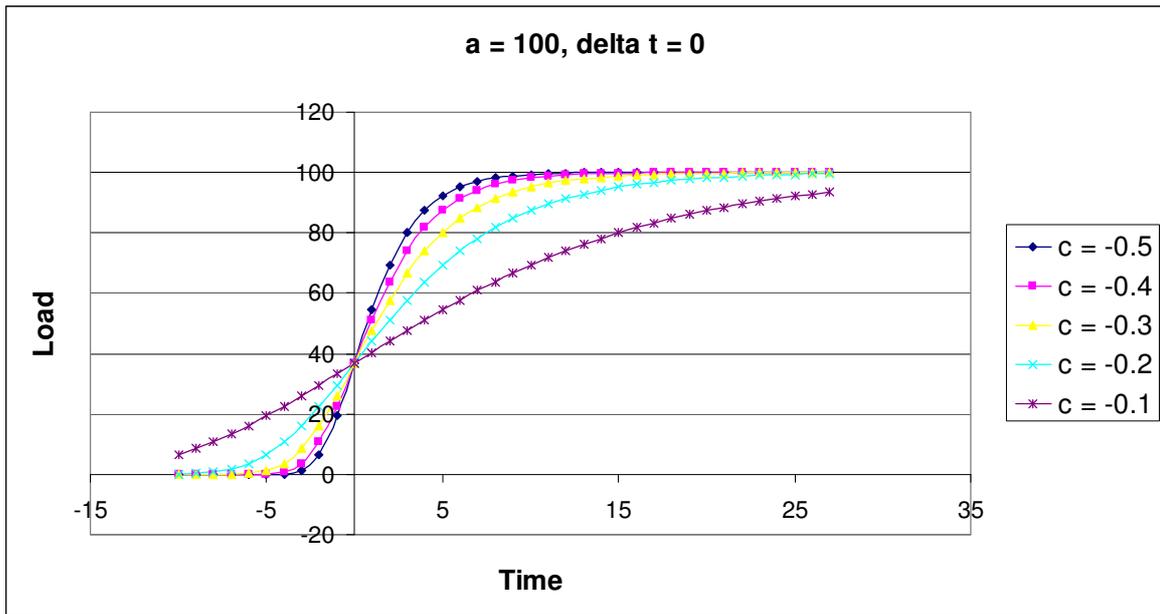


Figure 2.16 S-curve shapes affected by slope parameter

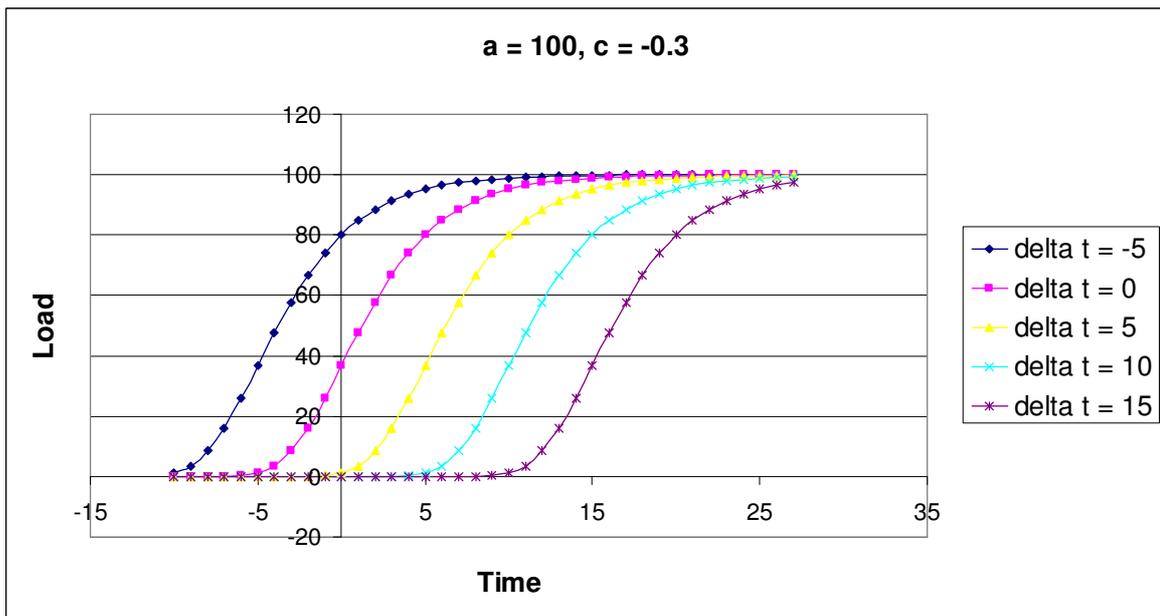


Figure 2.17 S-curve shapes affected by ramp up time parameter

2.6.2 Problem formulation

Bottom-up aggregation involves two elements:

- 1) Aggregate the historical load and HYL of the neighborhood small areas or sub-regions to obtain the historical load and HYL of the sub-region in the upper level.
- 2) Fit the historical load data of a small area or sub-region given the HYL using S-curve, which can be modeled as an optimization problem:

$$\begin{aligned} \text{Min:} \quad & \sqrt{\frac{1}{H} \sum_{t=1}^H \left(a e^{-e^{c(t-\Delta t)}} - h(t) \right)^2} \\ \text{s.t.} \quad & c \leq 0, \end{aligned} \quad (2.12)$$

where H is the length of historical load, and $h(t)$ is the historical load of the t th year.

This step is to tune the slope and shift the curve horizontally to find the best match with the historical load.

2.7 Top-Down Allocation: Multi-objective Optimization

Compared with the decision making process introduced in section 1.2, bottom-up aggregation simulates regional planners' forecasting process, while top-down allocation is like distributing the corporation forecast load growth to sub-regions in lower levels, and even down to the small areas in the bottom level (Figure 2.18). There are several goals during this process:

- 1) Generated S-curves should match the correspondent historical load.
- 2) The sum of future forecasted loads of neighborhood small areas or sub-regions in a given lower level should match the future forecasted load of the correspondent sub-region in the upper level.

- 3) Since the development period of regions with the same size is similar, the slope of small areas of sub-regions in the same level should be close.
- 4) The load in the farthest forecast year should be close to HYL.

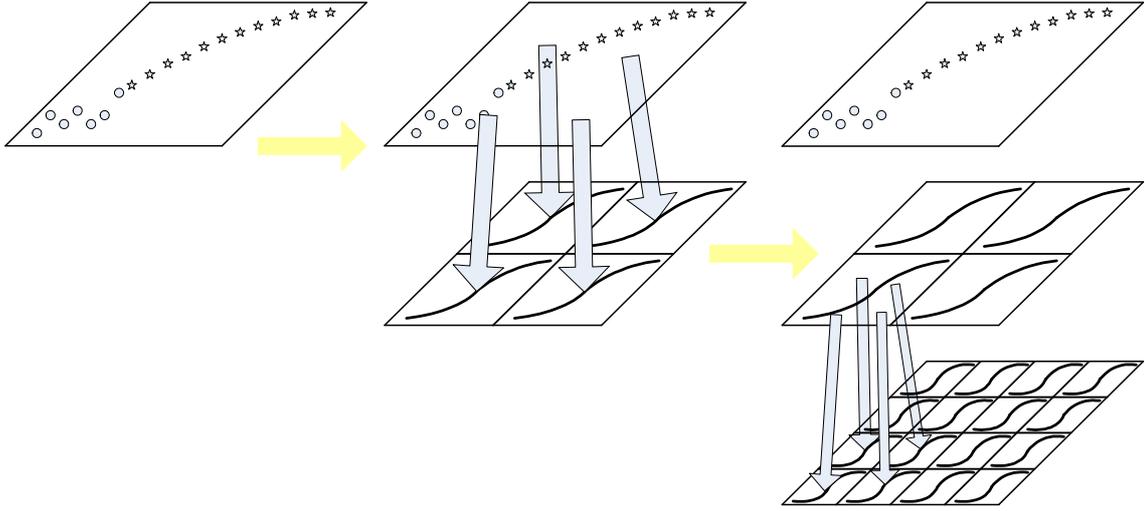


Figure 2.18 Top-down allocation

The corporation forecast load of the future years will be used in the top level to drive top-down allocation. To achieve the above goals, a multi-objective optimization approach is applied, where the first two are put into objectives, and the latter two are put into constraints:

- 1) Minimize historical load mismatch F_i :

$$F_i = \sqrt{\frac{1}{H} \sum_{t=1}^H \left(a_i e^{-e^{c_i}(t-\Delta t_i)} - h_i(t) \right)^2}, \quad i = 1, 2, \dots, K \quad (2.13)$$

where K is the number of neighbors in one neighborhood in a given level.

- 2) Minimize future load mismatch G :

$$G = \sqrt{\frac{1}{N-H} \sum_{t=H+1}^N \left(\sum_{i=1}^K (a_i e^{-e^{c_i}(t-\Delta t_i)}) - f(t) \right)^2} \quad (2.14)$$

where N is the total number of years including the historical years, and $f(t)$ is the forecasted load of correspondent sub-region in the upper level.

3) Slope boundary: $C_{\min} \leq c_i \leq C_{\max}$, where C_{\max} and C_{\min} represent the upper bound and lower bound of the slope.

4) Ramp up time boundary: $\Delta t_{\min} \leq \Delta t_i \leq \Delta t_{\max}$.

To sum up, each allocation unit can be formulated as:

$$\begin{aligned} \text{Min:} & \quad M_1 \sum_{i=1}^K F_i + M_2 G \\ \text{s.t.} & \quad C_{\min} \leq c_i \leq C_{\max}, \\ & \quad \Delta t_{\min} \leq \Delta t_i \leq \Delta t_{\max}, \end{aligned}$$

where M_1 and M_2 are the penalty factors for historical load mismatch and upper level future forecasted load mismatch respectively. In practice, M_1 is set to be much larger than M_2 to make the resulting forecast fit the historical load.

3 Human-Machine Co-construct Intelligence

3.1 Motivation

Although the automated computer program is convenient for users, the results lack common judgment. For instance, it may produce a higher load density of residential customers than industrial customers, which is intuitively incorrect to a distribution planner. Since there are many such rules based on planners' experience, it's neither efficient nor feasible to input all of them as constraints into the program. As an enhancement, a human expert is integrated into the problem solving loop to provide heuristics and insights to correct or confirm the results from the computer. The iterative calibration process, in which human and machine work together and negotiate with each other to come up with a solution, is named as human machine co-construct intelligence (HMCCI). Figure 3.1 shows a semi-automated program structure with the HMCCI framework implemented in the HYL module and the overall procedure illustrated.

Through the further development and use of practical analytical multi-objective planning techniques, analysts can begin to interact with all participants in the planning and management process and can enlighten any who would argue that electric resources policy evaluation and analyses should not be political. Analysts, utility company managers and planners have to work in a political environment. They need to understand the process of decision-making, what information is most useful to that process, and how it can best be presented. Knowledge of these facts in a particular planning situation might largely dictate

the particular approach to objective identification and quantification and to plan selection that is most appropriate.

The method deemed most appropriate for a particular situation will depend not only on the physical scale of the situation itself but also on the decision-makers, the decision-making process, and the responsibilities accepted by the analysts, the participants and the decision-makers.

Finally, because that the decisions being made at the current time are only part of a sequence of decisions that will be made on into the future. No one can predict with certainty what future generations will consider as being important or what they will want to do, but spending some time trying to guess is not an idle exercise. It pays to plan ahead, as best one can, and ask ourselves if the decisions being considered today will be those we think our descendants would have wanted us to make. This kind of thinking gets us into issues of adaptive management and sustainability

In the HYL module, distribution planners can critique the HYL density of each customer class based on their planning experience, e.g., the load density of industrial land should be much higher than that of residential land. Planners can also revise the land use data to his/her knowledge of the place he is familiar with. As an overall procedure, planners can add scenarios by adjusting the land use data and horizon year load of some particular small areas to meet the expected load growth. In the following two sections of this chapter, methods of using HMCCI to determine HYL and added scenarios will be discussed.

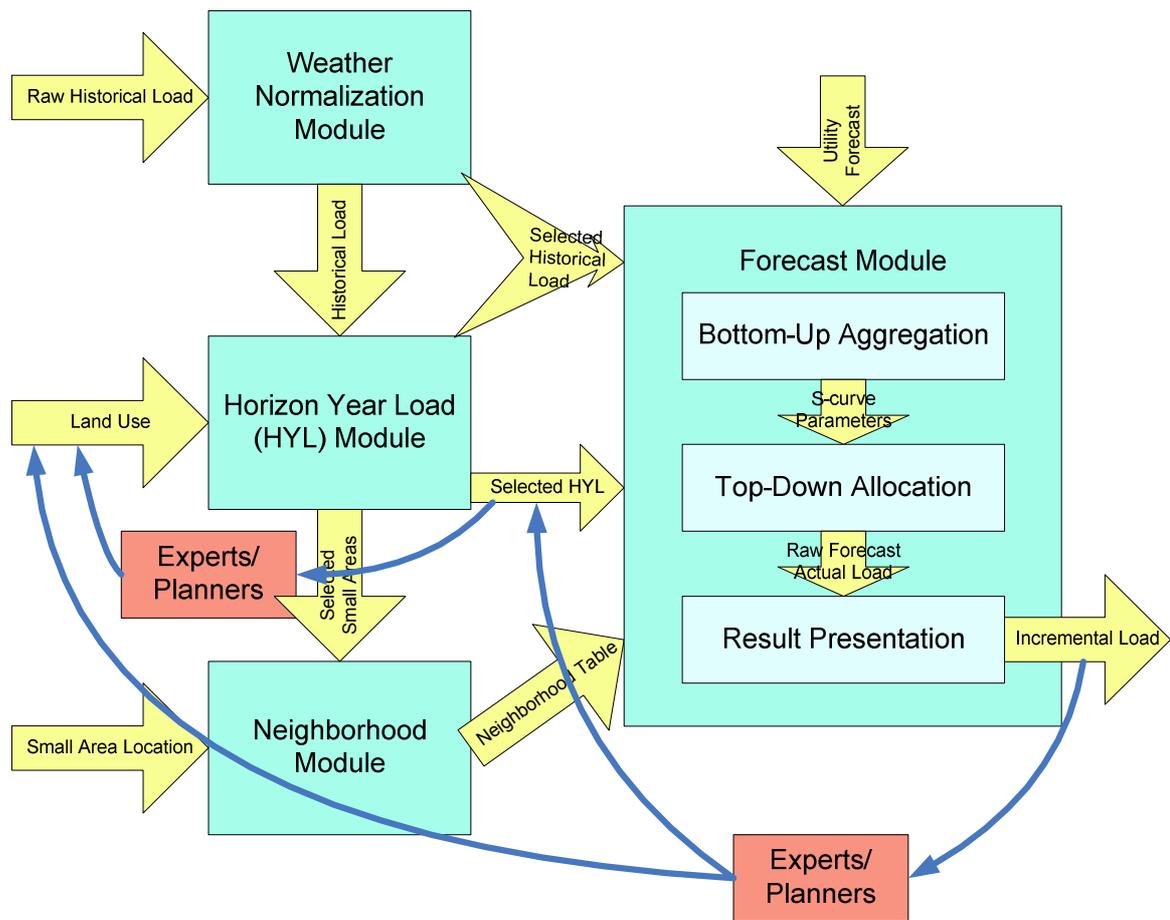


Figure 3.1 A semi-automated hierarchical trending method with HMCCI framework

3.2 Determining Load Densities Using Greedy HMCCI

3.2.1 Greedy Selection Procedure

As discussed in Section 2.4, load densities are calculated as intermediate results using current land use information and base year load, and HYL can be determined based on these load densities and future land use information. An optimization problem is formulated to calculate the load densities. Table 3.1 shows the resulting load densities when different amounts (from 100% to 99.0%) of the small areas are taken into consideration. Notice that all of these densities are equally reasonable to a computer running the optimization problem formulated in Section 2.4. However, an experienced distribution planner can understand the extensive physical meanings about these numbers and help the computer to select a more suitable set of load densities, or critique the current solutions to pursue better ones.

Table 3.1 Load densities for selection procedure

Land Use Information				Load Density (kW/acre)			
Index	Land Use Type	Current Acreage	Future Acreage	100.0%	99.8%	99.5%	99.0%
1	Industrial / Business	2490	4820	45.57	37.26	31.24	28.55
2	Extractive	1060	394	0.00	0.00	0.00	0.00
3	Commercial (Retail & Services)	4025	6204	34.30	39.29	37.60	40.41
4	Institutional / Government	2791	2733	45.89	26.71	24.04	22.25
5	Rural Residential	358	5553	31.54	4.55	9.92	11.74
6	Low Density Residential	16696	17446	2.64	4.81	6.26	6.58
7	Medium Density Residential	894	1607	36.62	35.49	31.21	32.84
8	High Density Residential	2525	3137	30.78	26.93	25.86	25.77
9	Communication / Utilities	695	673	13.53	15.68	15.89	4.83
10	Transportation	15372	16174	10.45	8.80	6.08	4.56
11	Agriculture / Vacant	66699	47679	0.00	0.00	0.00	0.00
12	Parks / Outdoor Recreation	5697	6802	0.00	0.27	0.65	0.96
13	Natural / Woodland / Water / Other	47490	50689	0.00	0.00	0.00	0.00

A greedy algorithm tries to find a local minimum at each stage of the problem with the hope of reaching a global optimum at the end. Although most of the time greedy algorithms fail to find the global optimum, they often produce a reasonably good approximation of the global optimum within a short time. Various greedy algorithms have been developed to solve NP-complete problems. However, due to the partial information utilized in each stage, computerized greedy algorithms may even reach the worst solution sometimes. This is definitely undesirable to some critical applications including load forecasting, because a bad forecast may lead to a disaster in a city and create much damage to the society. In this section, a human expert is put into the loop of the greedy strategy to have an overall control of the results in order to avoid worst case scenarios, or to even improve the results from the purely computerized program.

A greedy selection procedure can be described as following:

- 1) Find out the agreed upon observations among all the candidate solution sets. Highlight the observations using the color green, and exclude them from the future decision making.
- 2) Rank the significance of the remaining observations from more significant to less significant. If all of the remaining observations are similarly significant, just place them in any sequence.
- 3) Starting from the first observation in the queue, find out the solution set with the most disagreement, if there is one, among all the solution sets, and mark the observation in this solution set using the color red. If there is no disagreement on the observation,

move to the next one, until all the observations have been checked or there is only one solution set with no red mark. The remaining one or more solution sets without red mark will be the preferred choices.

The key argument here is that the good enough subset and the selected subset. Informally, a good enough subset is a subset of the search space in which the members satisfy some planning criteria set forth by the decision maker. Oftentimes, a good enough subset is easy to specify but difficult to obtain by the machine intelligence. In contrast, the selected subset is a subset in which the members are picked out by the human intelligence using certain evaluation analogies or metaphors as the outcome for the planning process. Every optimization problem can, in principle, be conceived as the goal of matching a selected subset with the good enough subset. Take Table 3.1 as an example:

- 1) All the solution sets agree that land use types #1, #11, and #13 have zero load densities, which are highlighted in green.
- 2) Suppose the remaining land use types are equally significant, so they are placed as they are numbered.
- 3) The load density of land use type #4 calculated by 100% small areas is marked as red, because it is significantly higher than the others.
- 4) The load density of land use type #5 calculated by 100% small areas is marked as red, because it is significantly higher than the others.
- 5) The load density of land use type #5 calculated by 99.8% small areas is marked as red, because it is significantly lower than the others.

- 6) The load density of land use type #9 calculated by 99% small areas is marked as red, because it is significantly lower than the others.
- 7) Therefore, load densities calculated by 99.5% small areas are selected as a reasonable solution set for further calculation of HYL.

3.2.2 Greedy Critique Procedure

Although the selection procedure allows a planner to select the best solution set in his/her mind among all the solution sets provided by the computer, it may still fail when none of the computer-provided solution sets are reasonable or when part of the solution sets have considerable drawbacks. A critique procedure can overcome such a failed scenario:

- 1) Starting with a solution set, find out the agreement between the planner and the computer. Mark the corresponding variables using the color green, make them as constants, and exclude them from the optimizations.
- 2) Among the remaining variables, find the most disagreed on variable, change its value, mark it as yellow, make it as a constant, and exclude it from the optimizations.
- 3) Tune the remaining variables to minimize the mismatches. If all of the resulting densities are colored by green or yellow, use them to compute HYL. Otherwise, go back to step (1).

Since in each time the planner and the computer will at least agree on one observation of all the finite many observations, the process will converge into a solution within finite many steps.

Table 3.2 shows a mock critique process starting with the calculation of 99.5% small areas, where the agreed upon values are highlighted in green and the manually modified values are highlighted in yellow:

- 1) The planner agrees with the load densities of land use type #2, #11, #12, and #13.
- 2) The planner changes the load density of transportation from 6.08kW/acre to 2.00kW/acre.
- 3) Among the resulting load densities, the planner agrees on land use type #1, #7, #8, and #9.
- 4) The planner changes the load density of the rural residential customers from 11.56kW/acre to 7.64kW/acre.
- 5) The planner agrees with none of the remaining resulting load densities.
- 6) The planner changes the load density of the commercial customers from 46kW/acre to 38kW/acre.
- 7) The planner is satisfied with all of the load densities. The load density results can be used in the HYL calculation.

Planners and managers working toward improving the performance of, or the solution sets provided by, these complex systems must identify and evaluate alternative plans and management or operating policies, comparing their predicted performance with desired goals or objectives. This is achieved through the greedy critique procedure.

Table 3.2 Using critique procedure to calculate load densities

Index	Land Use Type	Step 1	Step 2	Step 3	Step 4	Step 5	Step 6	Step 7
1	Industrial / Business	31.24	31.24	29.59	29.59	29.59	29.59	29.59
2	Extractive	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3	Commercial (Retail & Services)	37.60	37.60	44.79	44.79	46.00	38.00	38.00
4	Institutional / Government	24.04	24.04	19.97	19.97	18.02	18.02	19.20
5	Rural Residential	9.92	9.92	11.56	7.64	7.64	7.64	7.64
6	Low Density Residential	6.26	6.26	7.64	7.64	7.73	7.73	7.73
7	Medium Density Residential	31.21	31.21	34.88	34.88	34.88	34.88	34.88
8	High Density Residential	25.86	25.86	27.85	27.85	27.85	27.85	27.85
9	Communication / Utilities	15.89	15.89	17.05	17.05	17.05	17.05	17.05
10	Transportation	6.08	2.00	2.00	2.00	2.00	2.00	2.00
11	Agriculture / Vacant	0.00	0.00	0.00	0.00	0.00	0.00	0.00
12	Parks / Outdoor Recreation	0.65	0.65	0.65	0.65	0.65	0.65	0.65
13	Natural / Woodland / Water / Other	0.00	0.00	0.00	0.00	0.00	0.00	0.00

3.3 Customized Scenarios

Although well-tuned load densities can provide overall matching with the current load of most of the small areas, in a few small areas, which are treated as outliers, there may be some local rules that differ a lot from the calculated load densities. Even among the small areas used in the optimization problem, the variance of mismatches is still very large due to the local information that a computer program doesn't have. Therefore, such information has to come from the local planners.

Furthermore, because of the multitude of factors that determine demand, perfect forecasting of economic development and resulting demands is a utopian dream. Future demands are often dependent on future scenarios. An electric load demand scenario is a logical combination of basic parameters of the economy. An understanding of the functioning of the socio-economic system, based on the human expert knowledge developed through the assessment of past and present trends should be used to formulate a limited number of consistent scenarios. Therefore, integration of planners' local knowledge is essential for a useful forecast.

3.3.1 Revising Horizon Year Load

As discussed in the last chapter, HYL can directly affect the forecast. Thus, a planner can manipulate the forecast through revising HYL. For example, a local data center is going to gradually increase the number of computers during the next several years, which will apparently increase the future load consumption. However, this information will not be

reflected by land use information, because land use is the same as before. Therefore, the computer program is not able to indentify such growth. In this case, the planner can simply increase the HYL to include the load growth introduced by an increased number of computers.

Since HYL is calculated through land use data and base year load, land use data can affect the HYL as well. It is possible that errors exist in the land use data, which will lead to a wrong forecast if the error is significant. By looking at the forecast maps, a planner can tell by intuition whether the forecast is reasonable or not. For example, the forecast map may show that there is some load growth on a lake, which is a water body that has a zero load density. When checking the future land use data, the planner finds that some residential land use is misplaced on the lake. In this case, the planner can correct the land use data by erasing the residential land on the map.

3.3.2 Adding New Business Information

Back to the example of the data center extension example, if the increase in number of computers is going to take effect at the beginning of the following year, it is not suitable to model this by increasing the HYL, because there is suppose to be a jump in the middle of the original smooth S-curve. In this case, the planner can model the extension as a new business by adding the same land use that is equal to the original data center to the corresponding small areas. The extra load growth will be calculated separately and added to the top-down forecast results.

Redevelopment can be modeled as the combination of revising land use data and adding new business as well. Sometimes, old residential land may be redeveloped as commercial land. If future land use does not include such a redevelopment plan, the planner can subtract the residential land use from the future land use data, and add new commercial customers as new business to reflect the corresponding load growth.

4 Implementation and Results

The whole algorithm has been implemented in Microsoft Office Excel 2003 Visual Basic for Applications with Solver add-in. This chapter is devoted to introducing the implementation, Graphical User Interface (GUI), and related results.

4.1 Initialization

Some basic information is necessary to initialize the tool, which is put into the “Initialization” worksheet as shown in Figure 4.1.

In this worksheet, a planner needs to input the following information:

- 1) Names of the utility company and the planner. These two names will not be used in the program, but are for documentation purpose.
- 2) Temporal information that includes the base year, length of historical data, and forecast range. The first two will be used to generate the header of “Load History” worksheet (Figure 2.7).
- 3) Spatial information. The planner should input the length of a bottom-level small area, on which the size of the small area in both square footage and acreage is calculated automatically. The planner should also fill in the number of land use types which will be used to generate the headers of “Current Land Use” (Figure 2.5) and “Future Land Use” worksheets (Figure 2.6).

- 4) Corporation forecast. The planner can use a common percentage as the approximate corporation forecast, or specify the load growth of each year.

The screenshot shows a spreadsheet interface for a utility planner. The data is organized as follows:

- Utility Planner:** Madison Gas & Electric
- Temporal Information:**
 - Base Year: 2007
 - History Length: 7 (yr)
 - Forecast Range: 20 (yr)
- Spatial Information:**
 - Length: 1500 (ft)
 - Square Footage: 2250000 (ft²)
 - Acreage: 51.66 (Acre)
 - Number of Land Use Types: 13
- Corporation Forecast:**
 - Use a Common Annual Load Growth: 1.430%
 - Specify Load Growth By Year

Forecast Year	Projected Growth
2008	1.430%
2009	1.430%
2010	1.430%
2011	1.430%
2012	1.430%
2013	1.430%
2014	1.430%
2015	1.430%
2016	1.430%
2017	1.430%
2018	1.430%
2019	1.430%
2020	1.430%
2021	1.430%
2022	1.430%
2023	1.430%
2024	1.430%
2025	1.430%
2026	1.430%
2027	1.430%

Navigation buttons: Overview, Initialization, Load History, Current Land Use, Future Land Use, Weath.

Figure 4.1 "Initialization" worksheet

After completing the above information, the planner can click the "Next Step" button to generate the headers of "Load History", "Current Land Use" and "Future Land Use"

worksheets. Then the planner should fill in the three tables following the pre-defined format. The “Clear Data” button is used to reset the whole tool. By clicking this button, all the data input by the user and calculated by the program will be cleared.

4.2 Weather Normalization

Weather normalization is conducted after inputting the historical load, and current and future land use information. “Weather Normalization” is a stand-alone worksheet as shown in Figure 4.2. The data for calculation has been entered and hidden to the user.

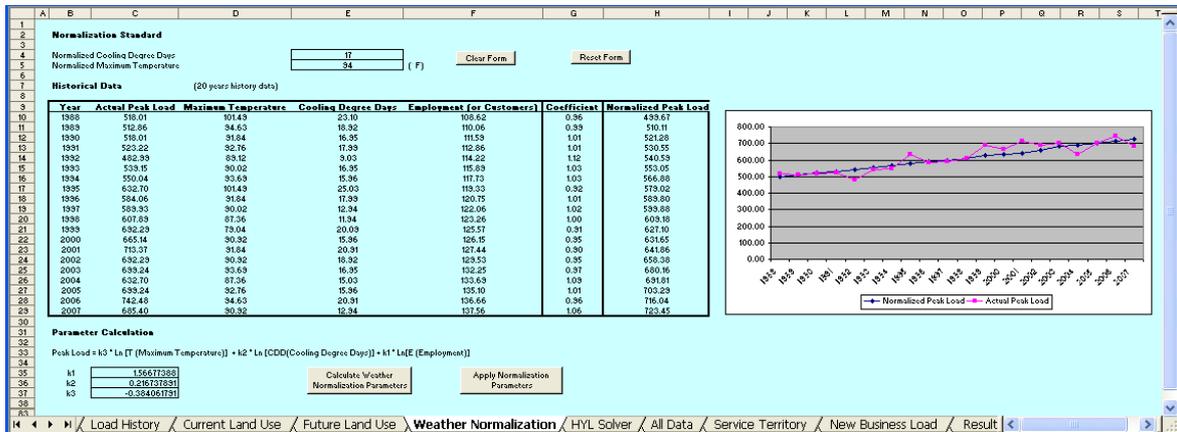


Figure 4.2 “Weather Normalization” worksheet

As introduced in Section 2.2.2, weather normalization applies a multivariate regression approach to calculate the coefficient for each historical year. There are four functional buttons in the “Weather Normalization” worksheet:

- 1) Clear Form: Delete all the data in this table.
- 2) Reset Form: Copy the pre-entered hidden data to the table.
- 3) Calculate Weather Normalization Parameters: Use multivariate regression to calculate the coefficient for each year.

- 4) Apply Weather Normalization Parameters: Multiply the raw historical load from the “Load History” worksheet by the corresponding coefficient and output the weather normalized load to the “All Data” worksheet as shown in Figure 4.3.

	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1															
2	SHAPE NUMBER	COORDINATE X	COORDINATE Y	2001	2002	2003	2004	2005	2006	2007	Horizon Year Load / Business	1	2	3	4
3	57530	2068500	402000	0	0	0	0	0	0	0	0	0	0	0	0
4	57531	2068500	403500	4.923404	4.50497	5.382981	5.752943	6.068078	3.204033	3.536886	12.41976825	0	0	0	0
5	57532	2068500	405000	0	0	0	0	0	0	0	0	0	0	0	0
6	57533	2068500	406500	0	0	0	0	0	0	0	0	0	0	0	0
7	57534	2068500	408000	0	0	0	0	0	0	0	0	0	0	0	0
8	57535	2068500	409500	0	0	0	0	0	0	0	0	0	0	0	0
9	57536	2068500	411000	0	0	0	0	0	0	0	0	0	0	0	0
10	57759	2070000	402000	0	0	0	0	0	0	0	0	0	0	0	0
11	57760	2070000	403500	7.59503	6.813297	10.57709	9.877694	8.512716	4.372203	5.710339	8.500039113	0	0	0	0
12	57761	2070000	405000	0	0	0	0	0	0	0	0.774271611	0	0	0	0
13	57762	2070000	406500	0	0	0	0	0	0	0	0	0	0	0	0
14	57763	2070000	408000	8.944183	7.375863	10.67152	10.52897	16.65787	8.702544	10.1378	19.13836885	0	0	0	0
15	57764	2070000	409500	3.692553	3.125366	4.249722	3.364929	3.260327	3.705323	3.04817	10.86431788	0	0	0	0
16	57765	2070000	411000	0	0	0	0	0	0	0	0	0	0	0	0
17	57987	2071500	400500	0	0	0	0	0	0	0	0	0	0	0	0
18	57988	2071500	402000	0	0	0	0	0	0	0	0	0	0	0	0
19	57989	2071500	403500	1.631135	1.250146	1.983204	2.062376	2.25539	1.01249	1.18123	3.174433408	0	0	0	0
20	57990	2071500	405000	11.0766	9.751141	15.299	16.60755	17.44198	9.195159	10.3068	16.91762784	0	0	0	0
21	57991	2071500	406500	0	0	0	0	0	0	0	0	0	0	0	0
22	57992	2071500	408000	13.2111	10.87627	11.33259	11.61443	12.43511	6.888824	2.285515	43.17187045	0	0	0	0
23	57993	2071500	409500	0.984681	21.12747	27.48154	27.24507	32.18551	9.89319	14.32082	28.78376827	0	0	0	0
24	58215	2073000	399000	0	0	0	0	0	0	0	0.001966583	0	0	0	0
25	58216	2073000	400500	0	0	0	0	0	0	0	0	0	0	0	0
26	58217	2073000	402000	0	0	0	0	0	0	0	0	0	0	0	0

Figure 4.3 “All Data” worksheet – weather normalized historical load

4.3 Horizon Year Load Solver

The HYL of each small area is calculated in the “HYL Solver” worksheet as shown in Figure 4.4. There are four functional buttons:

- 1) Initialize HYL Solver: Copy necessary data, such as the base year load, and the current and future land use data to the present worksheet. Then calculate a statistical summary (e.g., current and future acreage of each land use type) and set up formulas to prepare for Solver’s optimization. Finally copy current and future land use data to the “All Data” worksheet.
- 2) Calculate HYL: Run Solver twice to get the load densities as well as the HYL. All small areas are used for optimization the first time, and a user-determined percentage of small areas is used at the second time.
- 3) Apply HYL: Copy calculated HYL to the “All Data” worksheet.
- 4) Reset Form: Delete all the data and reset the form back to its original status.

Index	Type	Current Acreage	Future Acreage	Load Density (W/Wacre)
1	Industrial / Business	2490.10	4819.99	31.24
2	Extractive	1059.79	393.73	0.00
3	Commercial (Retail & Services)	4024.97	6203.92	37.60
4	Institutional / Government	2790.83	2732.71	24.04
5	Rural Residential	388.14	5552.98	9.92
6	Low Density Residential	16995.96	17445.62	6.26
7	Medium Density Residential	893.84	1607.49	31.21
8	High Density Residential	2525.33	3137.18	25.86
9	Communication / Utilities	695.49	679.49	15.99
10	Transportation	15371.63	16173.74	6.08
11	Agriculture / Vacant	66688.52	47679.26	0.00
12	Parks / Outdoor Recreation	5636.65	6902.14	0.65
13	Natural / Woodland / Water / Other	47489.80	50888.87	0.00

Shape Number	Horizon Year Load	Base Year Load
57530	0.00	0.00
57531	12.42	3.54
57532	0.00	0.00
57533	0.00	0.00
57534	0.00	0.00
57535	0.00	0.00
57536	0.00	0.00
57539	0.00	0.00
57560	9.59	6.71
57581	0.77	0.00
57582	0.00	0.00
57583	19.14	10.14
57584	10.86	3.05
57585	0.00	0.00
57587	0.00	0.00

Figure 4.4 “HYL Solver” worksheet

4.4 Service Territory Map

The “Service Territory” worksheet (Figure 4.5) has the following four functions:

- 1) Show service territory map and detailed small area information, such as historical load, HYL, and current and future land use.
- 2) Allow user to edit future land use and HYL.
- 3) Allow user to add new business or development.
- 4) Allow user to select different accurate level to run engine.

The territory should often coincide with an administrative unit (region, district, county etc.), because the administrative system usually requires an analysis of the functioning of the resources within its administrative boundaries. The system boundaries however, depend on its physical characteristics. They include the administrative area, but may extend over a larger area, depending upon the physical boundaries.

As shown in Figure 4.5, the “Service Territory” worksheet is made of two blocks: the information board and the map. The map is drawn according to the coordinates of small areas, while the information board (Figure 4.6) includes the active functions and buttons.

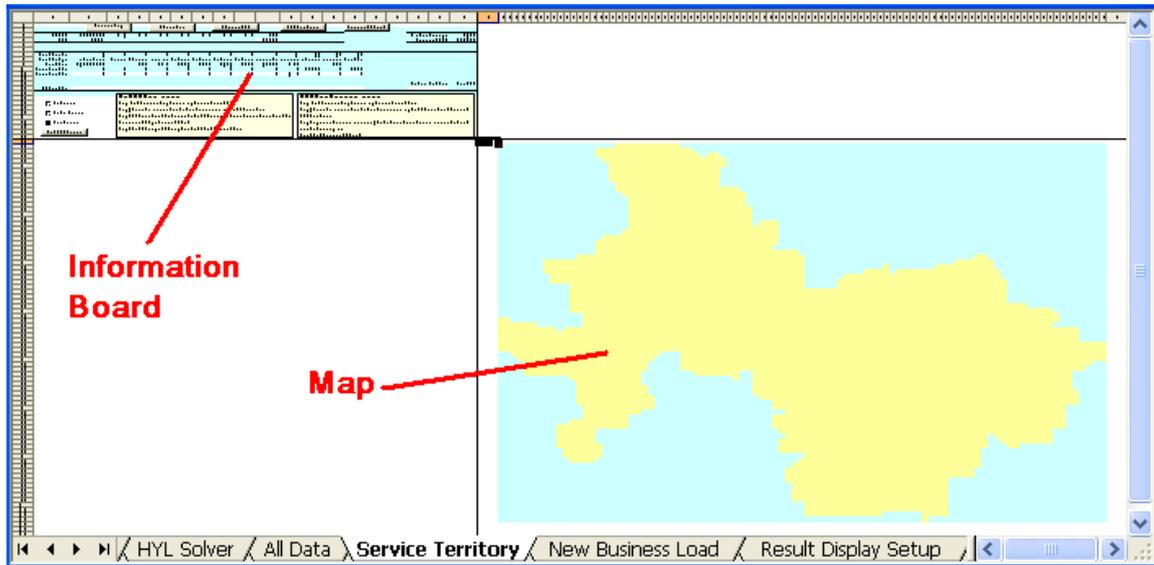


Figure 4.5 “Service Territory” worksheet

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	
1																				
2		Generate Map																		
3		AREA ID	SHAPE NUM	2001	2002	2003	2004	2005	2006	2007	HYL							Small Area Acreage	51.66	
4		856	63520	0	0	0	0	0	0	0	31.398549							Calculated HYL	31.39854911	
5																				
6																				
7		Land Use Index	1	2	3	4	5	6	7	8	9	10	11	12	13					
8		Land Use Type	Industrial / Business	Extractive	Services	Government	Residential	Other												
9		Load Density	31.23675744	0	37.80301	24.04133	9.921471	6.255429	31.20858	25.85582	15.885627	6.077923	0	0.852583	0					
10		Current Land Use	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	97.75%	0.00%	2.25%				
11		Future Land Use	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	10.00%	0.00%	0.00%	0.00%					
12																				
13																		Start Year	End Year	Delta HYL
14		NB Land Use																		0
15																				
16																				
17		<input type="radio"/> Fast Forecast	Modify HYL of a small area: Step 0: Click Generate Map if there is no map shown in the worksheet. Step 1: Select the small area to be edited on the service territory map. Click Choose Area. Step 2: Edit Future Land Use. Meanwhile the HYL correspondent to the new Future Land Use is shown in cell S4. Type the new HYL in J4. Step 3: Click Change HYL to update the HYL in All Data worksheet.																	
18		<input type="radio"/> Medium Forecast																		
19		<input type="radio"/> Slow Forecast																		
20		<input type="radio"/> Start INSITE Engine																		
21			Add New Business of a small area: Step 0: Click Generate Map if there is no map shown in the worksheet. Step 1: Select the small area to be edited on the service territory map. Click Choose Area. Then click Add New Business. Step 3: Input new business information in row 14, which includes land use information and start & end year of the development. Step 4: Click Calculate NB Load.																	
22																				
23																				
24																				

Figure 4.6 “Service Territory” worksheet – information board

There are five buttons on the information board:

- 1) Generate Map: Draw service territory map based on small area coordinates in “All Data” worksheet.
- 2) Choose Area: Show the detailed information of the selected small area on the service territory map.

- 3) Change HYL: A planner can change future land use data and HYL which are the cells filled with the color white. The modified HYL will be copied to the “All Data” worksheet to update the old one.
- 4) Add New Business: This row contains cells for new business land use, start year, and end year that are shown as white cells on the board. A planner can fill in the blanks with the corresponding new business information.
- 5) Calculate NB load: A new business load will be fit into an S-curve based on the planner’s input, and then put into the “New Business Load” worksheet.

Other than the five buttons, there are three check boxes on the information board as well. These boxes are used to select a forecast threshold, which is used to determine whether a small area’s load is considered as growth. If the mismatch between the HYL and the base year load is above the threshold, then the small area’s load can be considered as growth, and vice versa. Table 4.1 shows the approximate time under different forecast speed settings when the program is running on a computer with Intel Core 2 Duo 2.2GHz CPU and 2G RAM.

Table 4.1 Computer run time under different forecast speed settings

Check Box	Threshold (kW)	Amount of small areas	Time (min)
Fast Forecast	10	1024	50
Medium Forecast	5	1280	60
Slow Forecast	0	3460	110

Notice that the faster forecast has less accurate results than the slower one, because the threshold of the faster forecast is higher than the slower ones. Increasing the load growth threshold is one approach to reduce run time, while another more significant approach is to

reduce the resolution of small areas, e.g., from 51.66 acres to 206.64 acres. Both of them will sacrifice the accuracy of the forecast.

If a planner decides to delete some of the areas to increase the efficiency of the estimates, he or she of course needs some kind of procedure or criterion that indicates which areas best to delete. In reality, most decisions are made non-hierarchical, neither strictly bottom-up nor strictly top-down, as radically transformational, involving the development of partial and interim solutions which may ultimately play no role in the final decision. The process intrinsically involves the discovery of new insight.

4.5 Engine

4.5.1 Bottom-Up Aggregation

The small areas' historical load and the HYL are aggregated from bottom level to top level, and the data of each level are stored in a different workbook called "Level id" ("id" is replaced by the level index) as shown in Figure 4.7, where the first column is the index of the small area, the last column is HYL, and the middle columns are the load of historical years.

	A	B	C	D	E	F	G	H	I
1	1	8.944183	7.375863	10.67152	10.52897	16.65787	8.702544	10.13718	12.04089
2	2	13.21113	10.87627	11.33259	11.61443	12.43511	6.888824	6.285515	14.09999
3	3	0.984681	21.12747	27.48154	27.24507	32.18551	9.893193	14.32082	17.38255
4	4	16.57546	13.18904	26.53715	25.18269	27.64072	14.50343	17.37837	28.11416
5	5	27.40695	23.56526	31.8257	33.64929	32.75787	17.97121	21.39927	24.88806
6	6	0	7.250848	11.80478	12.59135	20.76343	11.61053	13.99714	31.69256
7	7	6.881394	8.389148	7.686442	10.74613	8.698717	9.906247	12.99541	15.01347
8	8	16.3203	18.0327	19.47434	19.22038	23.84748	27.01779	22.82805	24.8055
9	9	19.81287	19.90793	14.57918	20.29825	22.93167	21.65521	33.48798	36.01733
10	10	0	0	0	0	0	0	0	6.562421
11	11	41.34863	32.7438	38.40165	30.38695	33.12492	18.20816	25.85593	30.1091
12	12	34.31937	30.21757	34.63006	35.39323	29.613	13.3469	14.32584	16.97337
13	13	0	0	2.133425	2.329743	2.029091	1.877606	1.955296	4.00227
14	14	33.21674	28.95446	30.40131	31.90048	27.24295	16.43174	17.83283	20.02169
15	15	10.88847	10.29924	8.114635	7.101931	10.88175	6.379929	7.807898	10.50142
16	16	38.68078	35.0947	31.98008	33.36916	37.71916	43.71395	38.08834	46.00774
17	17	8.652958	8.084879	13.38504	13.12648	15.22174	17.38652	17.25118	19.42842
18	18	9.837161	12.90973	10.56886	12.83566	18.25698	14.64177	19.37793	22.43495
19	19	0	0	0	0	0	0	0	13.48622
20	20	25.43759	46.1304	103.9462	101.4473	98.62414	60.70812	69.26448	94.74017
21	21	62.77447	45.79862	62.70553	53.44557	40.3444	23.3674	25.7391	36.92159
22	22	7.982757	8.387303	9.682468	7.571663	8.90861	8.641172	8.483495	10.66876
23	23	17.47337	19.94314	22.26421	24.52701	25.74554	23.69132	21.33986	23.99599
24	24	0	0	0	0	6.319837	7.729213	7.526715	9.885844
25	25	0	0	0	0	0	0	0	3.186269

Figure 4.7 Temporary results of bottom-up aggregation

The bottom-up S-curve fitting is implemented in a single "Bottom-Up Solver" worksheet as shown in Figure 4.8. For each small area, historical data are copied onto this worksheet to calculate the S-curve parameters. These parameters will be stored in temporary worksheets call "Curve id" ("id" is replaced by the level index).

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1																
2		a	delta t	c					1	2	3	4	5	6	7	
3	1	644299.2	-2.84341	-0.058656					288165.2	301201.3	314062.1	326719.9	339150.3	351331.5	363244.6	
4							728526.1		291023.4	301478.1	307786.9	328488.2	339650.3	351223.3	364263	
5									2858.164	276.8614	-6275.12	1768.259	500.0053	-108.204	1018.399	
6									2726.81							
7																

Figure 4.8 “Bottom-Up Solver” worksheet

4.5.2 Top-Down Allocation

The top-down allocation is also implemented in a single worksheet named “Top-Down Solver” as shown in Figure 4.9. The upper level forecast data is copied from the corresponding “Level id” worksheet, and the initial value of the S-curves parameters are copied from the “Curve id” worksheet. After the Solver finishes the calculation for each sub-region’s load allocation, the forecast results will be copied to the corresponding “Level id” worksheet. All the worksheets generated by the engine, including the “Bottom-Up Solver” worksheet, “Top-Down Solver” worksheet, “Level id” worksheets, and “Curve id” worksheets (other than “level1” worksheet), will be deleted after the top-down allocation is finished.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
1	Region ID	a(H/L)	delta t	slope	MaxH	Mismatch H				Year #:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
2	1021	6153303	1144088	-0.38853	0	1544425				1	1.05E-09	8.79E-07	0.000104	0.000464	0.045899	0.398394	1.244092	3.484468	7.439568	12.90959	19.55843	26.49743	33.09886	38.9596	43.97709	48.05568
3	1022	594.8621	15	-0.24	0	0.431586				1	1.87E-10	8.7E-08	1.09E-05	0.000488	0.009709	0.102015	0.64891	2.781385	8.733053	21.50705	43.87611	76.25259	118.2053	166.8756	218.8725	270.9325
4	1023	795.0639	15	-0.24	5.284063	101.3191				1	2.37E-10	1.1E-07	1.38E-05	0.00062	0.012322	0.129468	0.823529	3.529818	11.0907	27.29451	65.42918	96.77311	150.0139	211.7811	277.7725	343.8393
5	1024	94.51068	15	-0.24	0	0.010891				1	2.97E-11	1.38E-08	1.73E-06	7.76E-05	0.000542	0.008206	0.103084	0.441638	1.388257	3.485338	6.33824	12.11539	18.7777	26.5093	34.76984	43.03943
6										0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7										0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8																										
9																										
10										SUM(TD)	1.8E-09	1.09E-06	0.00013	0.004849	0.063432	0.554681	2.819615	10.23749	28.65396	65.20868	125.6289	211.6375	320.095	444.1511	575.3335	705.8664
11																										
12																										
13																										
14										Region ID (H)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15										Objective	History	Forecast														
16										1	0.001															
17										166.9817	103.4057	63576.01														
18																										
19																										
20																										
21																										
22										Upper S:	0	0	4.741904	4.432441	4.615387	5.294063	4.180488	26.10227	54.44243	99.38278	162.6884	243.5061	338.8088	444.0148	554.0532	664.1645
23																										
24																										
25																										
26										Mismatch	1.05E-09	8.79E-07	0.000104	0.000464	0.045899	0.398394	1.244092									
27											1.87E-10	8.7E-08	1.09E-05	0.000488	0.009709	0.102015	0.64891									
28											2.37E-10	1.1E-07	-4.74189	-4.43282	-4.60306	-5.1646	-3.35686									
29											2.97E-11	1.38E-08	1.73E-06	7.76E-05	0.000542	0.008206	0.103084									
30											0	0	0	0	0	0	0									
31											0	0	0	0	0	0	0									
32																										
33																										
34										S-SUM	1.8E-09	1.09E-06	-4.74177	-4.43479	-4.54595	-4.73938	-1.36087	-15.8658	-25.7885	-34.1741	-37.0394	-31.8686	-18.7138	0.136338	21.34036	41.7084
35																										
36																										
37																										

Figure 4.9 “Top-Down Solver” worksheet

4.6 Forecast Results

4.6.1 Result Display Setup

A planner can select some particular years or all years within the forecast range to display forecast results in the “Result Display Setup” worksheet (Figure 4.10) by selecting one of the two check boxes.

FORECAST YEAR	Yes/No
2008	YES
2009	YES
2010	YES
2011	YES
2012	YES
2013	YES
2014	YES
2015	YES
2016	YES
2017	YES
2018	YES
2019	YES
2020	YES
2021	YES
2022	YES
2023	YES
2024	YES
2025	YES
2026	YES
2027	YES

Figure 4.10 “Result Display Setup” worksheet

4.6.2 Final Forecast Results

Final forecast results are presented in both data format and map format. Data format includes the incremental load (Figure 4.11) and the actual load (Figure 4.12). The incremental load is the load growth compared with the base year. If the growth is less than the threshold pre-determined in the “Service Territory” worksheet, it is treated as equal to the threshold. The actual load is the sum of the incremental load and the base year load.

1	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W		
2	AREA ID	SHAPE NUM	BASE YEAR				FORECAST YEAR																		
3	1	57530	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
4	2	57531	3.5368961	5.430327	6.149771	6.73572	7.207612	7.582384	7.877251	8.107598	8.286565	8.425033	8.531828	8.613994	8.677094	8.725484	8.762553	8.790926	8.818263	8.822223	8.841905	8.856595	8.859897	0	
5	3	57532	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
6	4	57533	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
7	5	57534	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
8	6	57535	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
9	7	57536	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
10	8	57759	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
11	9	57760	5.7103384	2.680445	2.706169	2.72586	2.740323	2.750439	2.761241	2.767966	2.773104	2.777028	2.780025	2.782313	2.784061	2.785395	2.786413	2.787191	2.787785	2.788238	2.788584	2.788848	2.789055	2.789215	
12	10	57761	0	0.036336	0.089323	0.163763	0.253440	0.346307	0.434729	0.511302	0.574595	0.624911	0.663635	0.693029	0.714962	0.731158	0.743027	0.75168	0.757962	0.76251	0.765793	0.768071	0.769691	0	
13	11	57762	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
14	12	57763	10.1371829	5.868287	6.558332	7.107785	7.53854	7.873904	8.134495	8.335959	8.491246	8.610668	8.702347	8.772634	8.826466	8.867664	8.899174	8.923263	8.941672	8.955738	8.966482	8.974688	8.980955	8.985997	
15	13	57764	3.04867002	3.965534	4.730657	5.370442	5.894031	6.31554	6.650634	6.94466	7.120277	7.281035	7.40511	7.500828	7.574467	7.633063	7.674454	7.707697	7.733143	7.752609	7.767492	7.778667	7.78559	0	
16	14	57765	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
17	15	57987	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
18	16	57988	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
19	17	57989	1.8112293	1.448278	1.568927	1.663916	1.73863	1.796383	1.842313	1.87738	1.904423	1.925228	1.941205	1.953456	1.962841	1.970024	1.975519	1.97972	1.98293	1.985333	1.987257	1.98889	1.989878	0	
20	18	57990	10.3306335	5.422398	5.630966	5.895959	6.058936	6.182335	6.277191	6.349969	6.405711	6.448441	6.48111	6.506093	6.525989	6.539782	6.550929	6.559444	6.566947	6.573913	6.574705	6.5776	6.57981	0	
21	19	57991	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
22	20	57992	6.28551493	14.63958	18.54838	22.01998	24.99331	27.47162	29.49487	31.12061	32.4122	33.42633	34.2193	34.83498	35.31139	35.67879	35.96144	36.17849	36.34493	36.47243	36.57001	36.64455	36.7012	0	
23	21	57993	14.3208197	12.16061	12.78847	13.24903	13.58497	13.82901	14.00577	14.13353	14.22574	14.29222	14.3401	14.37458	14.39939	14.41724	14.43008	14.43932	14.44596	14.45073	14.45416	14.45663	14.45841	0	
24	22	58216	0	0.001796	0.001839	0.001871	0.001895	0.001914	0.001927	0.001937	0.001945	0.00195	0.001958	0.00196	0.001962	0.001963	0.001964	0.001965	0.001965	0.001965	0.001965	0.001965	0.001965	0	
25	23	58216	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
26	24	58217	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
27	25	58218	0	0.005455	0.026124	0.086359	0.215134	0.43183	0.735046	1.103205	1.504084	1.905615	2.282875	2.620397	2.91267	3.154859	3.354442	3.515256	3.643187	3.743973	3.822783	3.884059	3.931497	0	
28	26	58219	4.25346275	3.366483	3.577188	3.741949	3.870054	3.963226	4.045746	4.104639	4.149877	4.184576	4.21161	4.231512	4.24708	4.258984	4.268082	4.275034	4.280345	4.284402	4.287499	4.289865	4.291671	0	
29	27	58220	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
30	28	58221	17.3783859	25.240693	30.27255	34.50231	37.98239	40.79531	43.03946	44.80967	46.18521	47.27426	48.10322	48.75501	49.25385	49.62478	49.82782	50.00239	50.15249	50.28588	50.40584	50.63233	50.63206	0	
31	29	58222	0	4.25E-05	4.35E-05	4.43E-05	4.49E-05	4.53E-05	4.56E-05	4.59E-05	4.61E-05	4.62E-05	4.63E-05	4.64E-05	4.64E-05	4.65E-05	0								
32	30	58441	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	

Figure 4.11 “Forecast Incremental Load” worksheet

1	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	
2	AREA ID	SHAPE NUM	BASE YEAR				FORECAST YEAR																	
3	1	57530	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
4	2	57531	3.5368961	9.867187	9.689631	10.27258	10.74447	11.1924	11.4141	11.6446	11.82342	11.96189	12.06869	12.15085	12.21935	12.26234	12.29941	12.32779	12.34949	12.36608	12.37876	12.38845	12.39686	0
5	3	57532	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	4	57533	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	5	57534	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	6	57535	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	7	57536	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	8	57759	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	9	57760	5.7103384	8.390784	8.416508	8.436769	8.451261	8.462776	8.47180	8.479305	8.483443	8.487367	8.490364	8.492852	8.494939	8.496724	8.498252	8.499572	8.499752	8.499833	8.499877	8.499923	8.499969	0
12	10	57761	0	0.036336	0.089323	0.163763	0.253440	0.346307	0.434729	0.511302	0.574595	0.624911	0.663635	0.693029	0.714962	0.731158	0.743027	0.75168	0.757962	0.76251	0.765793	0.768071	0.769691	0
13	11	57762	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	12	57763	10.1371829	16.00547	16.65701	17.24497	17.67534	18.01109	18.27168	18.47314	18.62843	18.74795	18.83963	18.90982	18.96365	19.00485	19.03636	19.06045	19.07886	19.09292	19.10366	19.11187	19.11814	0
15	13	57764	3.04867002	7.014204	7.775327	8.419102	8.942702	9.36421	9.699304	9.963167	10.16394	10.32557	10.45378	10.5495	10.62336	10.67973	10.72312	10.75637	10.78181	10.80128	10.81666	10.82754	10.83523	0
16	14	57765	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17	15	57987	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18	16	57988	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
19	17	57989	1.8112293	2.634041	2.75005	2.845039	2.919753	2.9781																

The “Forecast Map” worksheet is shown in Figure 4.13. In the upper-left corner, which is shown in larger detail Figure 4.14, the planner can select which year to look at and whether to view the incremental or actual load map. For instance, Figure 4.15 and Figure 4.16 show the incremental load map and the actual load map of 2009’s forecast results respectively. In both maps, the darker the color code is, the more load or load growth the small area has.

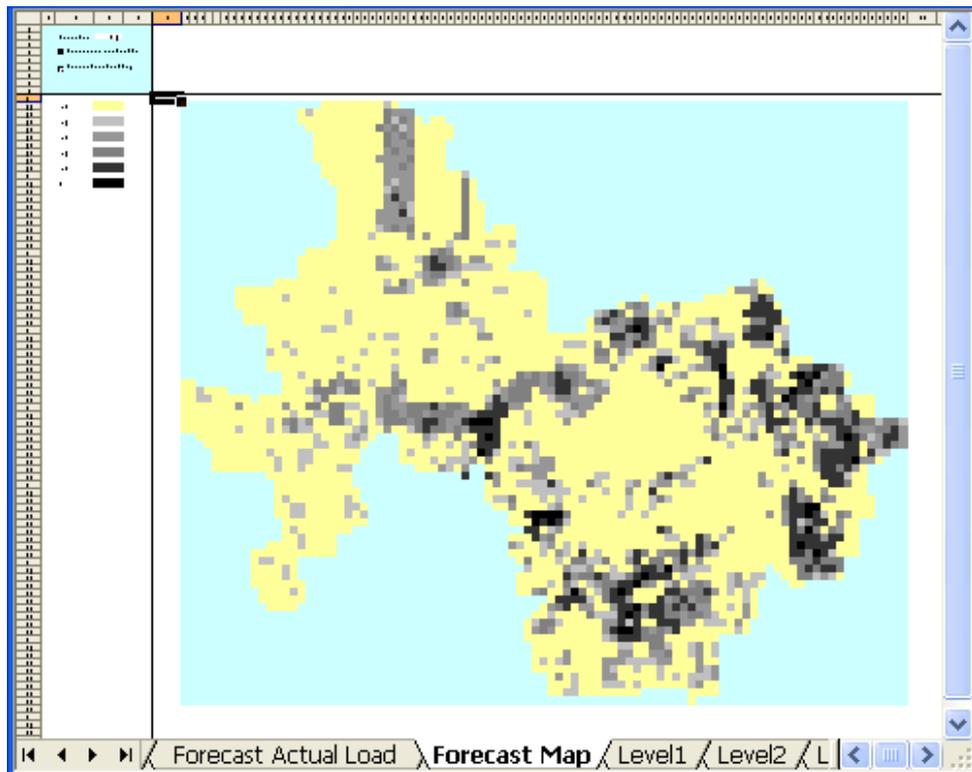


Figure 4.13 “Forecast Map” worksheet

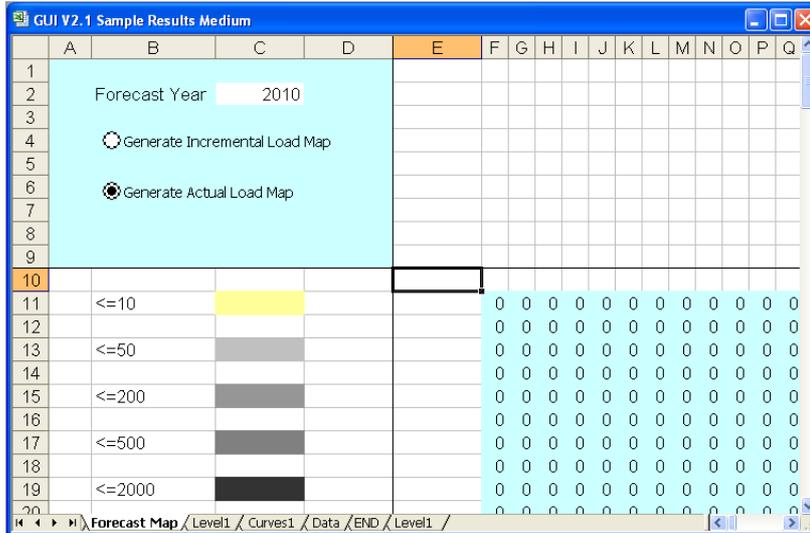


Figure 4.14 Forecast map selection

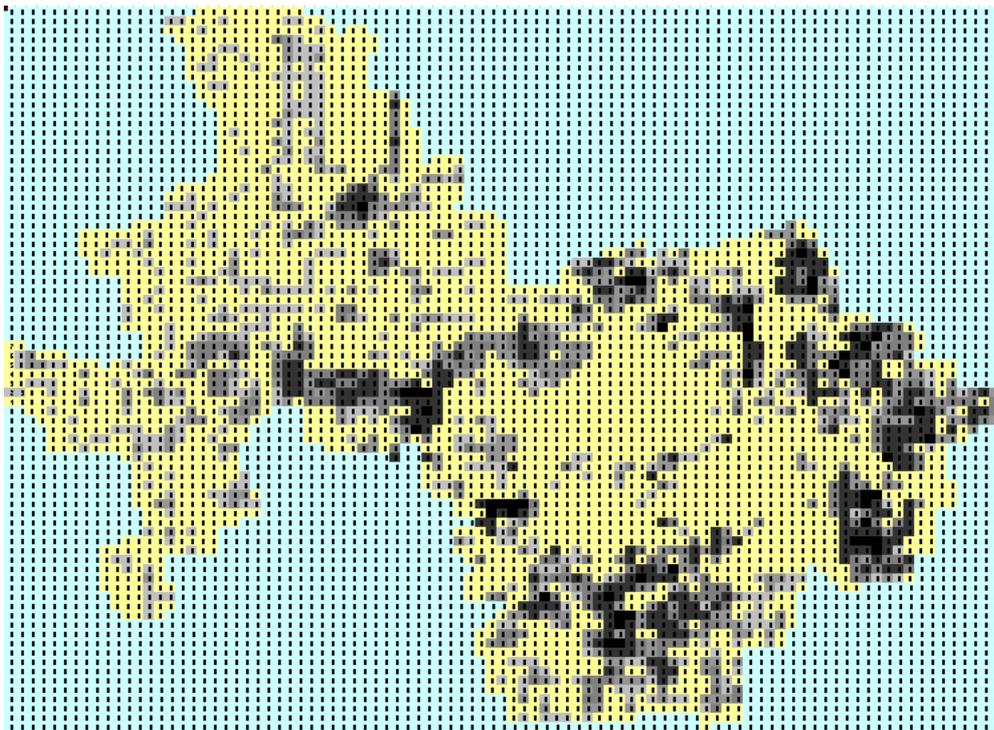


Figure 4.15 Forecast incremental load map (2010)

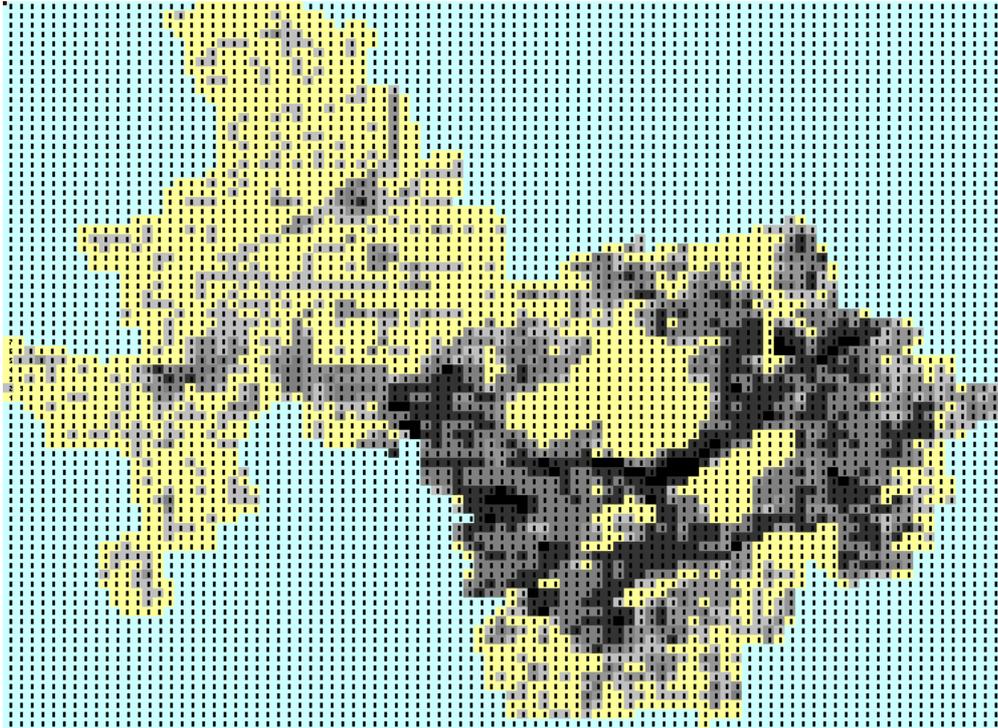


Figure 4.16 Forecast actual load map (2010)

5 Conclusion and Future Work

5.1 Conclusion

This thesis presents a tool for long-range spatial load forecasting. A hierarchical trending method has been investigated and implemented. An advanced feature, HMCCI, has been proposed to integrate human planners into the decision making process. The proposed method has been applied to several utilities and has received satisfactory results. Compared with most existing methods and tools in the literature and industry, the tool presented has the following advantages:

- 1) Easy to set up. The required user inputs can be extracted easily from most utilities' databases.
- 2) Short run time. It takes less than two hours to get the results for a service territory with 3400 small areas. Moreover, the run time can be further reduced by selecting a higher load growth threshold.
- 3) Short training period. It takes one business day to train the planners to understand the basic forecast concept, master the tool and produce forecast results.
- 4) Planner-friendly. HMCCI can allow more planners' involvement in order to get more reasonable results through several human-machine interaction cycles. Compared with other tools purely performed by the computer, this tool can bring in the planners' years of experience and knowledge of local development to guarantee the quality of the forecast.

Compared with most simulation methods, which are very popular in the current industry market, but require a tremendous investment of time and money to train planners, set up, and make a run, the proposed method is much more time and money efficient. However, the proposed method may not be as detailed or accurate in its results as simulation method is.

5.2 Future Work

Although the proposed tool has provided useful results for some utilities with limited budgets and resources for load forecasting, there are still several research topics that need to be further investigated in order to perfect the tool.

The current neighborhood table is built based on total number of small areas without the consideration of geographical information. The grouping of the nearest small areas or the small areas with similar land use types might be helpful to the forecast. The investigation of different grouping methods is expected to be one way to improve the current work.

In section 3.2.2, a simple greedy strategy is used to calculate load densities. Each time at least one type of land use is reduced, the whole procedure converges very fast. However, the accuracy of such an approach highly depends upon the skill of the planners. The proposed greedy strategy can be improved by adopting other more complicated greedy strategies, such as best first search, A* algorithm, etc. This will add more iterations to allow planners and computers to re-evaluate and revise what they've agreed on, so that even if the planners make a mistake, the computer has a chance to pick up on the error and give a warning.

References

- 1 Barakat, E.E.; Al-Rashed, S.A., "Long range peak demand forecasting under conditions of high growth," *IEEE Transactions on Power Systems*, vol.7, no.4, pp.1483-1486, Nov 1992
- 2 Barbounis, T.G.; Theocharis, J.B.; Alexiadis, M.C.; Dokopoulos, P.S., "Long-term wind speed and power forecasting using local recurrent neural network models," *IEEE Transaction on Energy Conversion*, vol.21, no.1, pp. 273-284, March 2006
- 3 Billinton, R.; Dange Huang, "Effects of Load Forecast Uncertainty on Bulk Electric System Reliability Evaluation," *IEEE Transactions on Power Systems*, vol.23, no.2, pp.418-425, May 2008
- 4 Blanchard, M.; Delorme, L.; Simard, C.; Nadeau, Y., "Experience with optimization software for distribution system planning," *IEEE Transactions on Power Systems*, vol.11, no.4, pp.1891-1898, Nov 1996
- 5 Yen-Ting Chao; Sheng-Ta Lee; Hong-Chan Chang; Tsai-Hsiang Chen, "An improvement project for distribution transformer load management in Taiwan," *IEEE Transactions on Power Systems*, vol.18, no.2, pp. 875-881, May 2003
- 6 Chen, C.S.; Hwang, J.C.; Tzeng, Y.M.; Huang, C.W.; Cho, M.Y., "Determination of customer load characteristics by load survey system at Taipower," *IEEE Transactions on Power Delivery*, vol.11, no.3, pp.1430-1436, Jul 1996
- 7 Chow, M.-Y.; Tram, H., "Methodology of urban re-development considerations in spatial load forecasting," *IEEE Transactions on Power Systems*, vol.12, no.2, pp.996-1001, May 1997
- 8 Chow, M.-Y.; Tram, H., "Application of fuzzy logic technology for spatial load forecasting," *IEEE Transactions on Power Systems*, vol.12, no.3, pp.1360-1366, Aug 1997
- 9 Mo-Yuen Chow; Jinxiang Zhu; Tram, H., "Application of fuzzy multi-objective decision making in spatial load forecasting," *IEEE Transactions on Power Systems*, vol.13, no.3, pp.1185-1190, Aug 1998
- 10 Christoforidis, M.; Aganagic, M.; Awobamise, B.; Tong, S.; Rahimi, A.F., "Long-term/mid-term resource optimization of a hydrodominant power system using interior point method," *IEEE Transactions on Power Systems*, vol.11, no.1, pp.287-294, Feb 1996

- 11 de la Torre, S.; Conejo, A.J.; Contreras, J., "Transmission Expansion Planning in Electricity Markets," *IEEE Transactions on Power Systems*, vol.23, no.1, pp.238-248, Feb. 2008
- 12 Gross, G.; Galiana, F.D., "Short-term load forecasting," *Proceedings of the IEEE*, vol.75, no.12, pp. 1558-1573, Dec. 1987
- 13 Hill, D.C.; Infield, D.G., "Modelled operation of the Shetland Islands power system comparing computational and human operators' load forecasts," *IEE Proceedings-Generation, Transmission and Distribution*, vol.142, no.6, pp.555-559, Nov 1995
- 14 Hung-Chih Wu; Chan-Nan Lu, "A data mining approach for spatial modeling in small area load forecast," *IEEE Transactions on Power Systems*, vol.17, no.2, pp.516-521, May 2002
- 15 Kandil, M.S.; El-Debeiky, S.M.; Hasanien, N.E., "Long-term load forecasting for fast developing utility using a knowledge-based expert system," *IEEE Transactions on Power Systems*, vol.17, no.2, pp.491-496, May 2002
- 16 Moghram, I.; Rahman, S., "Analysis and evaluation of five short-term load forecasting techniques," *IEEE Transactions on Power Systems*, vol.4, no.4, pp.1484-1491, Nov 1989
- 17 Palayanon, Visnu; Sinsawad, Arthorn; Juramongkorn, Borvorn; Willis, H. L.; Powell, R. W.; Tram, H. N., "Computerized Distribution Load Forecast for the City of Bangkok," *IEEE Transactions on Power Systems*, vol.2, no.1, pp.218-225, Feb. 1987
- 18 Powers, M.W., "The analytical Monte Carlo method for approximating the distribution of a plant's electrical generation," *IEEE Transaction on Energy Conversion*, vol.3, no.3, pp.433-439, Sep 1988
- 19 Quintana, V.H.; Temraz, H.K.; Hipel, K.W., "Two-stage power-system-distribution-planning algorithm," *IEE Proceedings-Generation, Transmission and Distribution*, vol.140, no.1, pp.17-29, Jan 1993
- 20 Ranaweera, D.K.; Karady, G.G.; Farmer, R.G., "Economic impact analysis of load forecasting," *IEEE Transactions on Power Systems*, vol.12, no.3, pp.1388-1392, Aug 1997
- 21 Smolleck, H.A.; Kim, K.C., "An interactive distribution load forecasting methodology for minicomputer use based upon a Markov-type process," *IEEE Transactions on Power Systems*, vol.3, no.1, pp.52-58, Feb 1988

- 22 Tram, H.N.; Willis, H.L.; Northcote-Green, J.E.D.; Brooks, C.L., "Load Forecasting Data and Data Base Development for Distribution Planning," *IEEE Transactions on Power Apparatus and Systems*, vol.PAS-102, no.11, pp.3660-3666, Nov. 1983
- 23 Willis, H.L., "Power Distribution Planning Reference Book, Second Edition, Revised and Expanded", Marcel Dekker, New York, 2002
- 24 Willis, H.L., "Load Forecasting for Distribution Planning-Error and Impact on Design," *IEEE Transactions on Power Apparatus and Systems*, vol.PAS-102, no.3, pp.675-686, March 1983
- 25 Willis, H.L.; Northcote-Green, J.E.D., "Spatial electric load forecasting: A tutorial review," *Proceedings of the IEEE*, vol.71, no.2, pp. 232-253, Feb. 1983
- 26 Willis, H.L.; Engel, M.V.; Buri, M.J., "Spatial load forecasting," *IEEE Computer Applications in Power*, vol.8, no.2, pp.40-43, Apr 1995
- 27 Willis, H.L.; Finley, L.A.; Buri, M.J., "Forecasting electric demand of distribution system planning in rural and sparsely populated regions," *IEEE Transactions on Power Systems*, vol.10, no.4, pp.2008-2013, Nov 1995
- 28 Willis, H.L.; Northcote-Green, J.E.D., "A Hierarchical Recursive Method for Substantially Improving Trending of Small Area Load Forecasts," *IEEE Transactions on Power Apparatus and Systems*, vol.PAS-101, no.6, pp.1776-1783, June 1982
- 29 Willis, H.L.; Northcote-Green, J.E.D., "Comparison Tests of Fourteen Distribution load Forecasting Methods," *IEEE Transactions on Power Apparatus and Systems*, vol.PAS-103, no.6, pp.1190-1197, June 1984
- 30 Willis, H.L.; Parks, T.W., "Fast Algorithms for Small Area Electric Load Forecasting," *IEEE Transactions on Power Apparatus and Systems*, vol.PAS-102, no.10, pp.3425-3432, Oct. 1983
- 31 Willis, H. Lee; Powell, R. W.; Tram, H. N., "Long-Range Distribution Planning with Load Forecast Uncertainty," *IEEE Transactions on Power Systems*, vol.2, no.3, pp.684-691, Aug. 1987
- 32 Willis, H.L.; Powell, R.W.; Wall, D.L., "Load Transfer Coupling Regression Curve Fitting for Distribution Load Forecasting," *IEEE Transactions on Power Apparatus and Systems*, vol.PAS-103, no.5, pp.1070-1076, May 1984
- 33 Willis, H.L.; Tram, H., "A Cluster Based V.A.I. Method for Distribution Load Forecasting," *IEEE Transactions on Power Apparatus and Systems*, vol.PAS-102, no.8, pp.2677-2684, Aug. 1983

- 34 Willis, H.L.; Tram, H.N., "Load Forecasting for Trnasmision Planning," *IEEE Transactions on Power Apparatus and Systems*, vol.PAS-103, no.3, pp.561-568, March 1984
- 35 Wu, L.; Shahidehpour, M.; Li, T., "Stochastic Security-Constrained Unit Commitment," *IEEE Transactions on Power Systems*, vol.22, no.2, pp.800-811, May 2007
- 36 Lei Wu; Shahidehpour, M.; Tao Li, "Cost of Reliability Analysis Based on Stochastic Unit Commitment," *IEEE Transactions on Power Systems*, vol.23, no.3, pp.1364-1374, Aug. 2008
- 37 Yeh, E.-C.; Tram, H., "Information integration in computerized distribution system planning," *IEEE Transactions on Power Systems*, vol.12, no.2, pp.1008-1013, May 1997