International Journal of Forecasting 31 (2015) 286–295

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Weather station selection for electric load forecasting

Tao Hong [a,*], Pu Wang [b], Laura White [c]

[a] University of North Carolina at Charlotte - Systems Engineering and Engineering Management, 9201 University City Blvd., Charlotte, NC 28223, United States

[b] SAS - R&D, Cary, NC, United States

[c] NCAEC - IT, Raleigh, NC, United States

ARTICLE INFO

Keywords:
Hierarchical load forecasting
Global energy forecasting competition
Short term load forecasting
Long term load forecasting
Weather station combination
Cross validation
Out-of-sample test
Greedy algorithm

ABSTRACT

Weather is a major driving factor of electricity demand. The selection of weather station(s) plays a vital role in electric load forecasting. Nevertheless, minimal research efforts have been devoted to weather station selection. In the smart grid era, hierarchical load forecasting, which provides load forecasts throughout the utility system hierarchy, is emerging as an important topic. Since there are many nodes to forecast in the hierarchy, it is no longer feasible for forecasting analysts to figure out the best weather stations for each node manually. A commonly used solution framework involves assigning the same number of weather stations to all nodes at the same level of the hierarchy. This framework was also adopted by all four of the winning teams of the Global Energy Forecasting Competition 2012 (GEFCom2012) in the hierarchical load forecasting track. In this paper, we propose a weather station selection framework to determine how many and which weather stations to use for a territory of interest. We also present a practical, transparent and reproducible implementation of the proposed framework. We demonstrate the application of the proposed approach to the forecasting of electricity at different levels in the hierarchies of two US utilities. One of them is a large US generation and transmission cooperative that has deployed the proposed framework. The other one is from GEFCom2012. In both case studies, we compare our unconstrained approach with four other alternatives based on the common practice mentioned above. We show that the forecasting accuracy can be improved by removing the constraint on the fixed number of weather stations.

1. Introduction

Electric load forecasting refers to the forecasting of electricity demand and energy a few minutes to a few decades ahead. As a fundamental business problem in the utility industry, load forecasting has extensive applications, such as power systems operations and planning, customer services, revenue management, energy trading, and so forth. Organizations in many sectors of the utility industry need load forecasts, such as the utilities themselves, regulatory commissions, retailers, and trading firms.

Hong (2014) offered an overview of the development of load forecasting practices over the past century. The load forecasting problem can be categorized roughly into two groups, short term load forecasting (STLF), where the forecast horizon is within two weeks (Hong, 2010), and long term load forecasting (LTLF), where the forecast horizon can be as long as 50 years or more (Hong, Wilson, & Xie, 2014). Many different techniques have been applied to STLF. Several representative ones include Artificial Neural Networks (Hippert, Pedreira, & Souza, 2001); multiple linear regression (Hong, 2010); time series analysis (Weron,

* Corresponding author.
E-mail address: tao.hong@uncc.edu (T. Hong).

http://dx.doi.org/10.1016/j.ijforecast.2014.07.001
0169-2070/

2006); semi-parametric additive models (Fan & Hyndman, 2012); and fuzzy regression (Hong & Wang, 2014). A recent literature review of STLF techniques and methodologies is provided by Hong (2010). The literature on LTLF is not as rich as that of STLF. Several papers have reported field-validated methodologies for LTLF, including a spatial load forecasting study by Hong (2008), a density load forecasting study at Australia Electricity Market Operator by Hyndman and Fan (2010), and a probabilistic load forecasting study at North Carolina Electric Membership Cooperative Corporation (NCEMC) by Hong, Wilson et al. (2014). Hong (2010) proposed an integrated load forecasting approach, which can use the same technique and data to develop both STLF and LTLF. This approach was reflected by Hong, Wilson et al. (2014), who applied multiple linear regression to STLF and LTLF. Discovered and practiced independently, this approach was also used by Hyndman and Fan (2010) and Fan and Hyndman (2012), who applied semi-parametric additive models to LTLF and STLF.

The massive deployment of smart grid technologies worldwide brings both opportunities and increased challenges to the load forecasting community. An emerging load forecasting approach, hierarchical load forecasting, involves forecasting the load at various voltage levels. Such forecasts offer the utilities more insights into power systems and customer usage patterns than the traditional high-voltage load forecasts. The availability of data varies depending upon the stage of smart meter deployment. In the US, even without any smart meters, most utilities have hourly load data down to distribution substations via SCADA systems. Those with smart meters deployed have hourly or sub-hourly load data down to the household level. Utilities need forecasts at low voltage levels so that they can perform distribution system operations better, such as circuit switching and load control. Even at the Independent Systems Operator level, leveraging the hierarchical load and weather information at the zonal level can help enhance the load forecasting accuracy of the aggregated loads (Lai & Hong, 2013). Hong, Pinson, and Fan (2014) proposed six challenges for hierarchical load forecasting. Two of them relate to ways of utilizing the hierarchical structure of the load, and of utilizing multiple weather stations spread across a large geographic area.

The literature on hierarchical load forecasting is limited, though there are a few major milestones in the area. Hong (2008) implemented a hierarchical trending method for spatial load forecasting at a medium-sized US utility, which involved fitting S curves for the 3000+ small areas and their aggregated levels through a constrained multi-objective optimization formulation. Fan, Methaprayoon, and Lee (2009) reported the results of a multi-region forecasting project at a Generation and Transmission (G&T) co-op. While the project aimed at high voltage load forecasting, the methodology was designed to look for the optimal combination of the regions to improve the forecasting accuracy. The study used the average of all weather stations. Hong, Wang, Pahwa, Gui, and Hsiang (2010) presented a simulation study showing the effects of temperature history data uncertainties on load forecasting accuracy. Lai and Hong (2013) reported an empirical hierarchical load forecasting case study based on ISO New England data, which included several ways of averaging weather stations and grouping loads.

The IEEE Working Group on Energy Forecasting organized the Global Energy Forecasting Competition (GEFCom2012), which included a track on hierarchical load forecasting. The competition problem was to backcast and forecast the loads of 20 zones and their sum using temperature information from 11 weather stations. Hong, Pinson et al. (2014) offered an overview of the methodologies used by 11 entries of the hierarchical load forecasting track. The four winning teams also published their solutions in the *International Journal of Forecasting* (Ben Taieb & Hyndman, 2014; Charlton & Singleton, 2014; Lloyd, 2014; Nedellec, Cugliari, & Goude, 2014).

Since the weather is a major driving factor of electricity demand, it is a common load forecasting practice today to include weather variables, such as temperature, humidity, wind speed and cloud cover, in the model. In the US, there are thousands of weather stations, each of which takes readings on an hourly or sub-hourly basis. However, many utilities still use only a few weather stations, in spite of operating over large territories. The aforementioned challenge in relation to weather station selection can be dissected into two questions: (1) how many weather stations should be used for a load zone, and (2) which weather station(s) should be used to forecast the load.

One of the characteristics of hierarchical load forecasting is the large number of nodes in the hierarchy. Because of this, it is not practical to investigate each node manually to assign the best weather stations. The four winning entries of GEFCom2012 tackled this challenge using different methods, but within the same framework. Charlton and Singleton (2014) built 11 models, one per weather station, and combined the forecasts from the five best-fit models. Lloyd (2014) used temperatures from all 11 weather stations. Nedellec et al. (2014) used a stepwise procedure based on a cross validation result to select one station for each zone. Ben Taieb and Hyndman (2014) used a testing week to select one station for each zone. The benchmark selected one station for each zone based on the best fit from 11 weather stations. As was reported by Hong, Pinson et al. (2014), all of these methods used a heuristically picked number (1, 3, 5 or all) of weather stations for each zone. In other words, the contestants assigned a number subjectively as the answer to the first question mentioned above, then tried to figure out which weather stations to use. This is a common solution framework in today's hierarchical load forecasting practice, where each node is assigned the same fixed number of weather stations. There has been no significant effort in the literature to identify the optimal number of weather stations for an individual load zone.

In this paper, we propose a weather station selection framework for electric load forecasting which aims to answer the two questions mentioned above simultaneously, namely how many and which weather stations to use for a territory of interest (Section 2). We also present a practical, transparent and reproducible implementation of the proposed framework, and illustrate it by forecasting the electricity usage in a supply area of NCEMC (Section 3). We then apply the proposed approach, which is not constrained by a fixed number of weather stations, to forecasting electricity at different levels in the hierarchies of