# Report Outline

Fabian

December 26, 2014

## 1 Introduction

A short introduction to the field of load forecasting and its relevance.
Description of the competition.

## 2 Review of related Work

Review previous work from Gefcom 2012. Distinguish between temperature and load prediction.
Review papers with related approaches including but not limited to those that can be found under the following link: http://blog.drhongtao.com/2014/08/recommended-papers-for-gefcom2014-contestants.html

## 3 Data

Description of the data made available for the competition

### 3.1 Temperature Data

The temperature data made available consists of 25 series of temperature data in Fahrenheit from 25 different weather statios dating from 01/01/2001 to 12/01/2011.
The Cross Correlation of the different temperature series with each other suggest that they can be explained to over 90% by the first series. [include correlation plots] If we assume a temperature of 60 degrees fahrenheit, that would allow for an error of maximum 6 degrees of fahrenheit or 3 degrees celsius. As we will see later this error is negligable given the inaccuracy of the temperature prediction.

## 3.2 Load Data

### 3.2.1 "Lag Analysis"

The effect of the lag on the Time Series Correlation Coefficient can be demonstrated using the Autocorrelation function acf() built into R stats. Here five plots showing the Autocorrelation Function for different maximum lags are displayed:

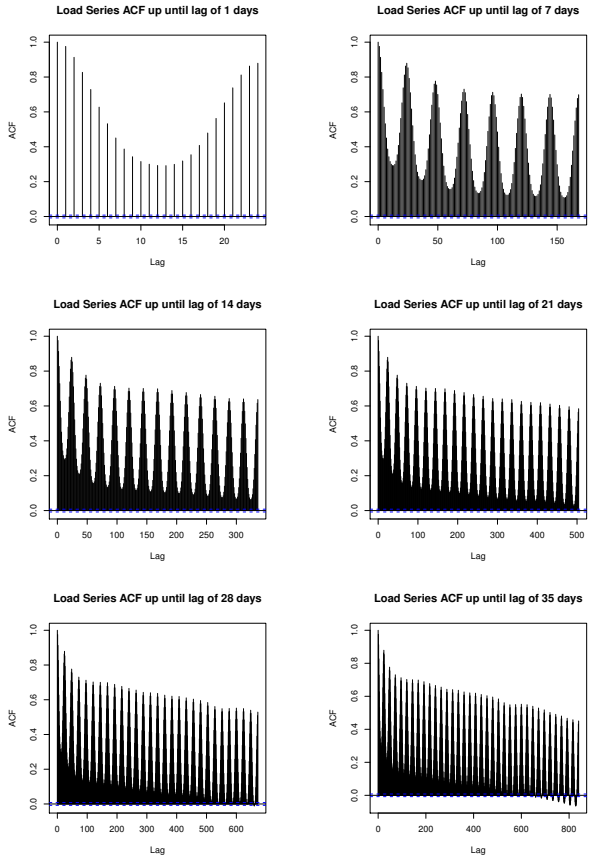


Figure 1: Plots of Autocorrelation Function estimates of hourly load data in Mega Watts for different maximum lags.

As can be seen the correlation diminishes exponentially up until a lag of 72h from a point in the time series, then stays more or less constant for up to 7-8 days whereafter it diminishes near linearly (up to a lag of 35 days).

### 3.3 Basic exploration with Time Series Analysis Methods

Autocorrelation, include decompositions? (included in "Data")

## 4 Feature 'Extraction'

Description of the features obtained from the data.

### 4.1 Calendar Features

hour, TOY vs. month

## 5 Models

Description of the Models (LM), GAM (more extensive), NN, RF

## 6 Analysis

What combination of features and models for temperature and load provide us with a good prediction accuracy with respect to Gefcom leaderboard?

### 6.1 Error Measures

Introduce error measures (RMSE, MAE, MAPE, PINBALL) and their differences here? Too late?

### 6.2 Temperature Modeling

#### 6.2.1 Data Processing

average temperature vs. principal component

#### 6.2.2 Effect on Load Prediction

Effect of temperature on load prediction evaluated using different methods: Mean over past years (yearly lag), LM, GAM, NN, RF vs. true temperature

#### 6.2.3 Evaluate Results of weekly vs. monthly Temperature Prediction

plot MAPE & PINBALL scores for different methods over load training + CV period in a 1x2 plot of the form:
monthly scores  weekly scores

### 6.3 Load Modeling

#### 6.3.1 Evaluate the Influence of the Lag on Load Prediction

set the basis by plotting MAPE & PINBALL scores by week over w1, w2, w3, w4; one curve for every month in CV
do this for every method
as well as a comparison of the best performing configuration of every method among each other

#### 6.3.2 Evaluate Performance for different Method Configurations

Use temp method that provides best score as shown in Temperature Modeling Section

Different GAM formulas:
plot MAPE & PINBALL scores for all GAM formulas over CV period in a 2x2 plot of the form:
monthly load with monthly temp  monthly load with weekly temp
weekly load with monhtly temp  weekly load with weekly temp

Different NN hidden units:
plot MAPE & PINBALL scores in 2x2 plot

Different RF ntrees:
plot MAPE & PINBALL scores in 2x2 plot

#### 6.3.3 Compare Performance of different Methods

choose best scoring configuration for every method and plot the results in one 2x2 plot

## 7 Conclusion

Draw conclusion based on the analysis done in the main part.