Load Forecasting: Oh Dear!

S. A. Soman
Yogesh K. Bichpuriya

PowerAnser Labs
Indian Institute of Technology Bombay, Mumbai India

May 26, 2016

Outline

1. PowerAnser Labs
2. Introduction to Forecasting
3. Short Term Load Forecasting
4. Long Term Load Forecast
5. Medium Term Load Forecasting
6. Linear Combination of Forecasts
7. Conclusion

PowerAnser Labs

- Joint initiative of IIT Bombay, TCS and TCE.
- Vision: Be a leader in developing knowledge based IT solutions for electric power systems
- web-services developed by PAL
  1. Load Flow Analysis (webLFA)
  2. Transmission System Cost/Loss Allocation (webNetUse)
  3. Optimal Power Flow (webOPF)
  4. Short Circuit Analysis (webSCA)
  5. Transmission Expansion Planning (webTEP)
  6. Short Term Forecasting (webSTF)
  7. Long Term Load Forecasting (webLTLF)
  8. Portfolio Optimization (webPO)
- Clients
  1. Tata Power Distribution Company Mumbai
  2. Central Electricity Regulatory Commission
  3. National Load Despatch Center

Introduction to Forecasting

1. Important analytics for many planning and operational decision making in
   - Electricity sector
   - Meteorology
   - Supply industry
   - Tourism sector
   - ...
2. A well studied branch (and yet evolving) in econometrics, statistics and operational research.
3. Restructuring of power systems and introduction of electricity markets have led to more accurate and robust forecasts of electrical load.
4. New forecasting problems are being introduced in electrical industry like wind forecasting, price forecasting etc.
5. Forecasting is even more important and challenging today - it stimulates research in modeling techniques and algorithm development.

Education: the path from cocky ignorance to miserable uncertainty. -*Mark Twain*

Types of Forecasts

A probability density forecast of the load can be defined as a probability distribution of all the realizations of the load under various scenarios of the predictor variables.

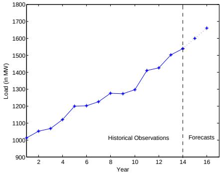


Figure: Point forecast

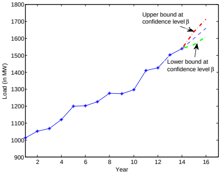


Figure: Interval forecast

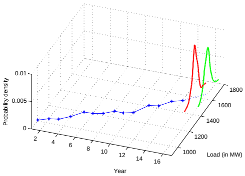


Figure: Probability density forecast

Introduction to Load forecasting problems

	Long term	**Medium term**	**Short term**
Time horizon	Year(s) ahead	Month(s) ahead	Day(s) ahead
Quantity of interest	Peak/average load in the forecasted year	Hourly peak/average load of a day during a forecast month	hourly/15-minutes load (load profile) of next day
Type of forecast	Probability density forecast	Probability density forecast	Point forecast
Load profile composition	Trend component	Seasonal and trend components	Seasonal component
Important exogenous variables	Economic, demographic and development indicators	Weather and economic indicators	Day type and weather
Applications	Portfolio management (resource planning) and network planning	Portfolio management (resource planning)	Day ahead trading and scheduling

Understanding Trend and Seasonal Components

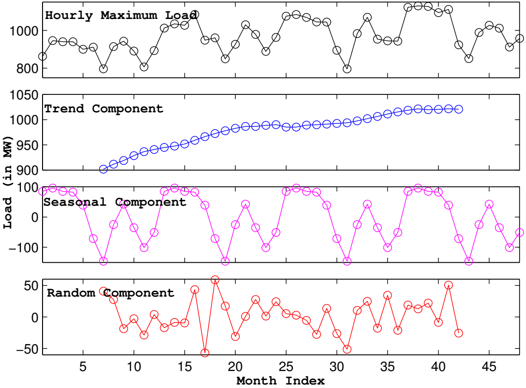
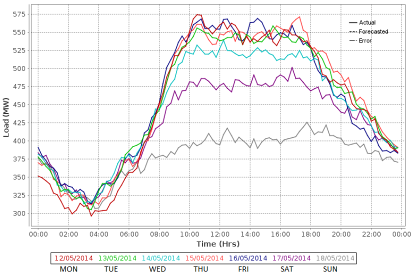
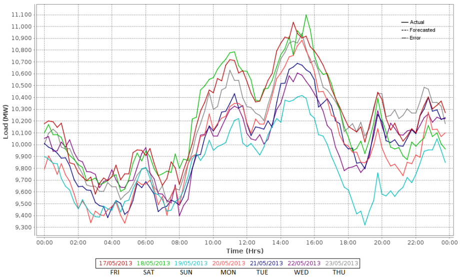


Figure: Monthly peak load of a distribution utility

Load Profiles of Distribution Companies

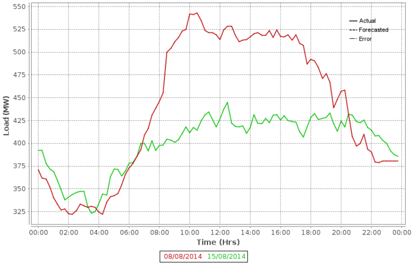


(a) Urban Distribustion Company

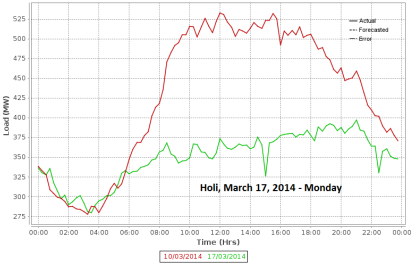


(b) State Distribution Company

Load Profiles on Holidays vs Normal Day

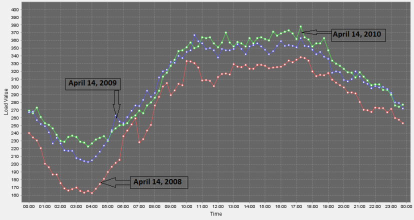


(c) Independence Day vs Normal Day

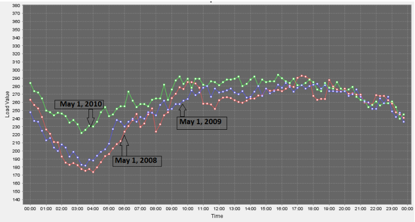


(d) Holi vs Normal Day

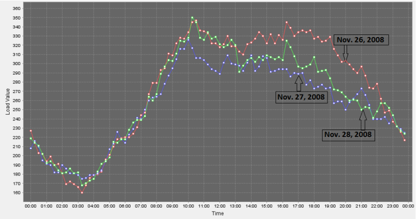
Load Profiles for Holidays/Calamities



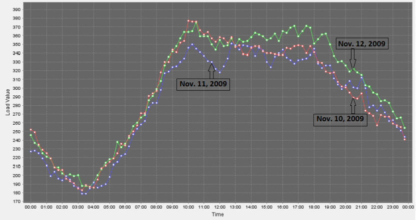
(e) Ambedkar Jayanti (Apr. 14)



(f) Maharashtra Day (May 1)



(g) Terror Attack (Nov. 26, 2008)



(h) Cyclone (Nov. 10, 2009)

Statistical Analysis of Data

- Mean-Variance analysis - Coefficient of Variation

$$CV = \frac{\sigma}{\mu}$$

- Check for Stationarity
- Correlation study
    - Similar days
    - Temperature Vs load
    - Humidity Vs load
    - Rainfall Vs load
    - $\cdots$
- Outlier/anomalous profiles detection
- $\cdots$

Forecasting Methods

- Regression Models
  - Similar day approach
  - · · ·
- Time Series Models
  - Autoregressive (AR)
  - Moving Average (MA)
  - Autoregressive Moving Average (ARMA)
  - Integrated Autoregressive Moving Average (ARIMA)
  - Generalized Autoregressive Conditional Heteroskedastic (GARCH)
- Evolutionary and Learning Models
  - Expert System (ES)
  - Genetic Algorithm (GA)
  - Particle Swarm Optimization (PSO)
  - Fuzzy Systems (FS)
  - Artificial Neural Network (ANN)
  - Support Vector Machine (SVM)
  - Self Organized Map (SOM)

Performance Analysis

- Mean Absolute Error (MAE)

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N} |L_i - F_i|$$

- Sum of Squared Error (SSE)

$$\text{SSE} = \sum_{i=1}^{N} (L_i - F_i)^2$$

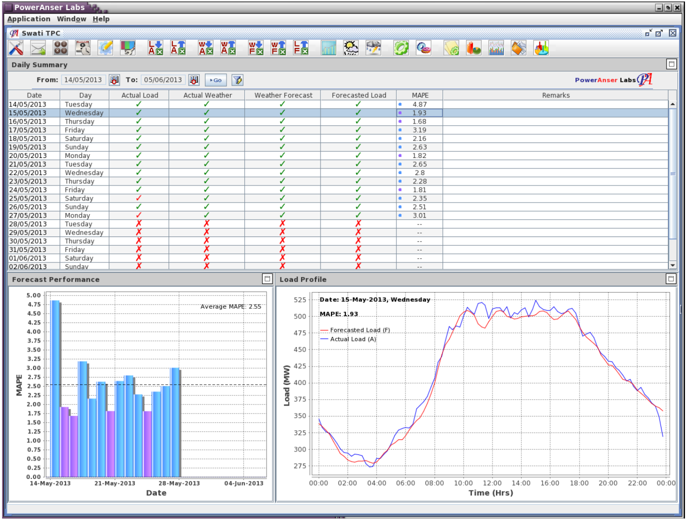
- Root Mean Squared Error (RMSE)

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} (L_i - F_i)^2}$$

- Mean Absolute Percentage Error(MAPE)

$$\text{MAPE}\ \% = \frac{1}{N}\sum_{i=1}^{N} \frac{|L_i - F_i|}{L_i} \times 100$$

webSTLF



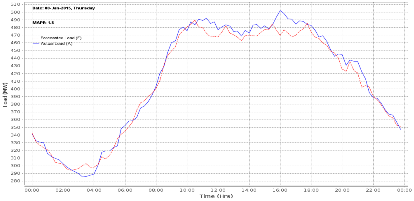
Date	Day	Actual Load	Actual Weather	Weather Forecast	Forecasted Load	MAPE	Remarks
14/05/2013	Tuesday	✓	✓	✓	✓	4.87	
15/05/2013	Wednesday	✓	✓	✓	✓	1.93	
16/05/2013	Thursday	✓	✓	✓	✓	1.68	
17/05/2013	Friday	✓	✓	✓	✓	3.19	
18/05/2013	Saturday	✓	✓	✓	✓	2.16	
19/05/2013	Sunday	✓	✓	✓	✓	2.63	
20/05/2013	Monday	✓	✓	✓	✓	1.82	
21/05/2013	Tuesday	✓	✓	✓	✓	2.65	
22/05/2013	Wednesday	✓	✓	✓	✓	2.8	
23/05/2013	Thursday	✓	✓	✓	✓	2.28	
24/05/2013	Friday	✓	✓	✓	✓	1.81	
25/05/2013	Saturday	✓	✓	✓	✓	2.35	
26/05/2013	Sunday	✓	✓	✓	✓	2.51	
27/05/2013	Monday	✓	✓	✓	✓	3.01	
28/05/2013	Tuesday	✗	✗	✗	✗	--	
29/05/2013	Wednesday	✗	✗	✗	✗	--	
30/05/2013	Thursday	✗	✗	✗	✗	--	
31/05/2013	Friday	✗	✗	✗	✗	--	
01/06/2013	Saturday	✗	✗	✗	✗	--	
02/06/2013	Sunday	✗	✗	✗	✗	--	

Challenges in STLF

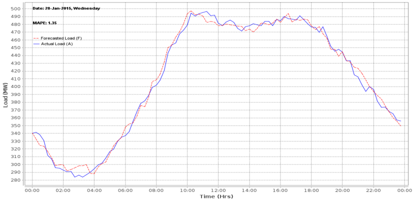
- Expectation of Users (e.g., Indian Utilities/Regulators expect around 2% MAPE).
  - Which forecasting model can achieve it, if at all !
  - We consider SDA, ANN and time series as three possibilities.
  - Modeling of unrestricted demand is a challenge.
- For STLF, one size fits all is not true.
  - Weather modeling for DISCOM based in New Delhi is going to be very different from DISCOM based in Mumbai.
  - Modelling of Holidays - Some are generic but others are region or state specific.
  - If a weekend is preceded or succeeded by a holiday, load profile could be significantly impacted.
- Presence of anomalous data
  - Anomalous data should not be used in forecasting.
  - If it is used in forecasting, forecast quality degrades.
  - Misclassification also degrades quality of forecast.

Yogesh K. Bichpuriya, Royden S. S. Fernandes and S. A. Soman, "Identification of anomalous load profile for short term load forecasting," PowerCon 2012, Auckland, Oct.2012.

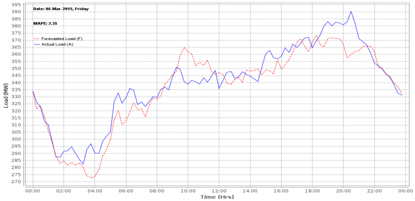
Performance of webSTLF on Weekdays and Holidays for an Urban Distribution Company



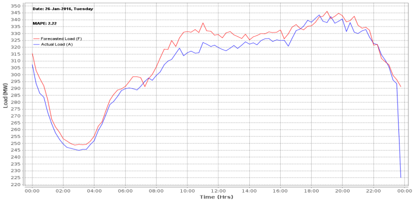
(i) Thursday, Jan 08, 2015 (MAPE=1.8%)



(j) Wednesday, Jan 28, 2015 (MAPE=1.35%)

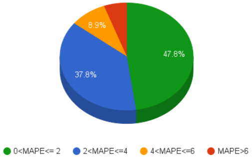


(k) Holi, Mar 06, 2015 (MAPE=2.35%)

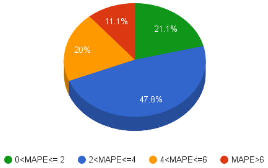


(l) Republic Day, Jan 26, 2016 (MAPE=2.22%)

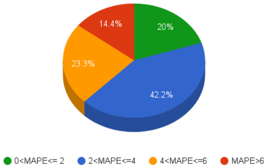
Performance for the period of three months



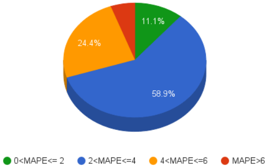
(m) State Discom 1



(n) State Discom 2



(o) State Discom 3



(p) Urban Discom 1

Probability Density Forecast of Yearly Peak Load

Problem description

- Long term (e.g., 5 years) forecast of yearly peak load of an electricity distribution company
- The load depends on explanatory variables like GDP, population, weather etc.
- Influential observations affect the estimation of trend
- Error in projection of explanatory variables may affect the accuracy of load forecast

Proposed solution

- Robust probability density forecast of yearly peak load
- Robustness against
  - influential observations
    - achieved by using a method akin to jackknifing
  - error in projection of explanatory variables
    - by considering probability density forecast of explanatory variables
  - discrepancies in estimation trend model
    - use of Alternating Conditional Expectation (ACE) [Breiman and Friedman, 1985] to discover trend and combination of probability density forecasts obtained using non-parametric model (ACE) and parametric models (linear, exponential)

Effect of Influential Observations

Table: Out-of-sample ex-post point forecasts

Year	Actual	Linear	Exponential	ACE
2007-08	1425.00	1412.84	1427.60	1439.90
2008-09	1501.00	1479.47	1523.25	1583.39
2009-10	1538.00	1528.52	1594.43	1657.60
2010-11	1675.00	1582.19	1677.85	1765.03
2011-12	1690.00	1646.43	1795.23	1830.98
MAPE		2.20%	2.35%	5.61%

Table: Out-of-sample ex-post point forecast after removing influential observation

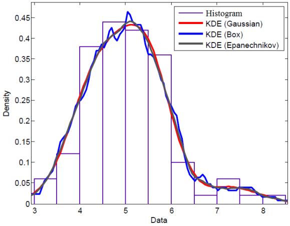
Year	Actual	Linear	Exponential	ACE
2007-08	1425.00	1317.93	1325.87	1316.15
2008-09	1501.00	1301.53	1326.12	1305.02
2009-10	1538.00	1311.38	1346.35	1310.97
2010-11	1675.00	1316.35	1364.05	1315.52
2011-12	1690.00	1345.73	1420.38	1331.70
MAPE		15.46%	13.12%	15.62%

Table: Out-of-sample ex-post point forecast with $N-1$ jackknife approach

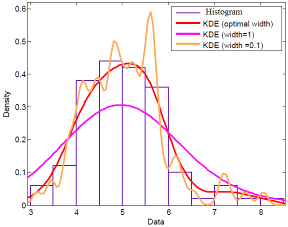
Year	Actual	Linear	Exponential	ACE
2007-08	1425.00	1413.11	1427.55	1425.19
2008-09	1501.00	1477.83	1519.34	1537.36
2009-10	1538.00	1526.65	1588.20	1592.38
2010-11	1675.00	1579.75	1669.51	1670.73
2011-12	1690.00	1645.12	1786.14	1710.53
MAPE		2.29%	2.14%	1.49%

Kernel Density Estimation

- Density estimation is the construction of an estimate of probability density function from the observed data [Silverman, 1986].
- *Parametric Approach* - assumed that the data are drawn from one of a known parametric family of distributions. Attributes of the distribution can be estimated using the data.
- *Non-parametric approach* does not require any assumption regarding distribution of the data. Histogram is a basic example of non-parametric density estimation method.
- Choice of Kernel includes: Epanechnikov,Gaussian, Box.
- Density estimated by KDE depends on the value of its smoothing parameter or bandwidth *bw*.
- If *bw* is small, end user can be misled by spurious results.
- Too large a bandwidth can mask salient features e.g., multimodal distribution may appear to be unimodal.



(q) KDE with kernel variation



(r) KDE with bandwidth variation

Alternating Conditional Expectation

- Consider a function

$$z = 5\sin(3x) + y^2. \quad (1)$$

- Problem is to estimate the function using given sample.
- Alternating Conditional Expectation (ACE) is a non-parametric approach to find the optimal transformation function for predictor and response variables[Breiman and Friedman, 1985].
- In ACE, we can represent the function as

$$\theta(z) = \phi_1(x) + \phi_2(y). \quad (2)$$

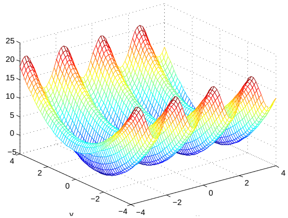


FIGURE: Plot of the function

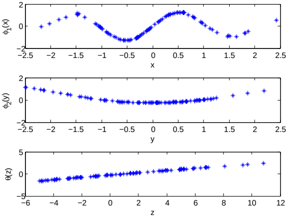


FIGURE: Plot of transformed function

Pseudo-code for Probability Density Forecast of Yearly Peak Load

```
Let N_s be the number of scenarios of explanatory variables and N be the total number of observations.
Given a trend model, the peak load density forecast for year t is to be calculated.
for s = 1 to N_s do
    Estimate the model parameters using all the data points.
    Given the scenario s of explanatory variables, compute forecast instance of the peak load.
    for i = 1 to N do
        Estimate the model parameters while deleting i^th observation from the data set.
        Compute a peak load instance using the estimated model.
    end for
    Calculate median of all forecasts instances for a given scenario.
end for
Estimate the density forecast using KDE.
```

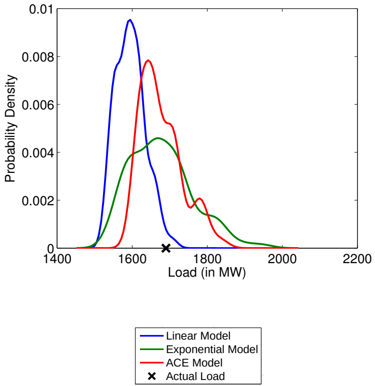
Yogesh K. Bichpuriya, S. A. Soman, A. Subramanyam, "Robust Probability Density Forecasts of Yearly Peak Load using Non-Parametric Model" in IEEE PES General Meeting 2016 *(accepted)*.

Case Study - Data Description

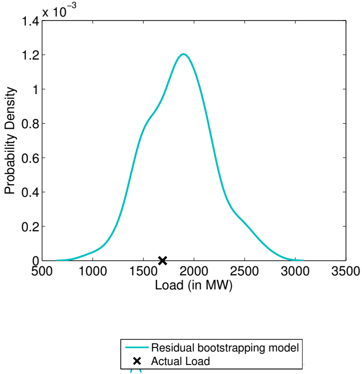
- Case study has been done on data set of an urban electricity distribution company in India.
- Historical load data for yearly peak load (in MW) was available from Financial Year (FY) 1994-95 to 2011-12.
- Set of explanatory variables - GDP, population and humidity.
- Data from FY 1994-95 to 2006-07 for estimating coefficients of the models
- Data from FY 2007-08 to 2011-12 for out-of-sample testing purpose
- Out-of-sample testing is done ex-ante.
  - In ex-ante case, we use projection of explanatory variables.

Probability density forecast of peak load in year 2011-12

Probability density forecasts of yearly peak load using the proposed approach with different trend models.



Probability density forecast using the residual bootstrapping method for peak load density forecast as suggested by [Veall, 1987].



Verification of Probability Density Forecast

Continuous Ranking Probability Score (CRPS)[Hersbach Hans, 2000]

$$CRPS(P, L) = \int_{-\infty}^{\infty} (P(y) - \mathbb{H}(y, L))^2 \, dy \quad (3)$$

where $P(y)$ be the cumulative density function of the forecast, L is the actual load. $\mathbb{H}(y, L)$ is the Heaviside function, defined as follows.

$$\mathbb{H}(y, L) = \begin{cases} 0 & \text{for} \quad y < L \\ 1 & \text{for} \quad y \geq L \end{cases} \quad (4)$$

For example, a point forecast can be considered as a probability density forecast by representing it as an impulse function as shown in the figure. It gives a lower bound on CRPS $=0$.

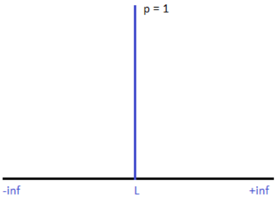


Figure: Probability density representation of a point forecast

Table: Comparison of CRPS for different trend models

Trend models	2007-08	2008-09	2009-10	2010-11	2011-12
Linear	20.43	47.93	34.92	124.43	72.50
Exponential	14.53	28.06	12.88	71.41	22.19
ACE	3.90	12.60	7.68	59.10	17.29
Residual Bootstrapping	12.53	33.22	37.30	67.08	107.05

Probability Density Forecast of Medium Term Load

Problem Description

- Forecast peak load (in MW/MVA) in a specified time interval (e.g., 10:00 hours-11:00 hours) during a month
- The peak load is aggregated at 11 kV of the distribution system which supplies electricity to residential, commercial and industrial consumers.
- We propose an application of a non-parametric method known as Alternating Conditional Expectation (ACE).

Proposed Model

$$\theta_h(L_h) = \phi_{(1,h)}(t) + \phi_{(2,h)}(T_h) + \phi_{(3,h)}(H_h) \tag{5}$$

where,

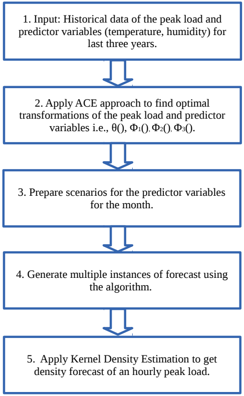
$$\begin{aligned}
L_h &= \text{peak load of hour } h, \quad (h = 1, 2, \cdots, 24) \\
t &= \text{Month index}, = 1, \cdots, N \;\; \text{(Typically, } N = 36) \\
T_h &= \text{Average temperature of hour } h \;\; (^\circ\text{C}) \\
H_h &= \text{Average humidity of hour } h \;\; (\%).
\end{aligned}$$

Yogesh K. Bichpuriya, S. A. Soman, and A. Subramanyam, "Probability Density Forecast of Hourly Peak Load of a Month using Non-parametric model," Power Systems Computation Conference (PSCC) , Wroclaw, Poland, August 2014.

Block diagram for the proposed approach

- $m_d(F)$ = number of days in a month for which forecast is required
- h = index for forecasted hour
- $j = 1, \cdots, 96$ be 15 minutes blocks in a day

for $h = 1$ to 24 **do**
  for $i = 1$ to $m_d(F)$ **do**
    for $j = 4(h-1)+1$ to $4h$ **do**
      Fetch forecasted values of $T_h(i,j)$ and $H_h(i,j)$.
      Using eqn. (5), obtain $L_h(i,j)$
    end for
  end for
  Estimate the density forecast using KDE.
end for



Case Study - Data Description

- The proposed approach is illustrated with a case study on the real life data of an urban distribution company in India.
- The historical load data of four years from May 2008 to April 2012 is considered.
- The weather data includes temperature and humidity.
- We consider 36 months (from May 2008 to April 2011) data for training i.e., to find optimal transformation of the predictor and response variables.
- Transformation function of the response variable i.e., the peak load is considered as linear.
- The optimal transformations of the response and predictor variables are obtained using ACE.
- Forecast is done six month ahead (for October 2011).

Six months ahead Ex-post out-of-sample density forecast of peak load

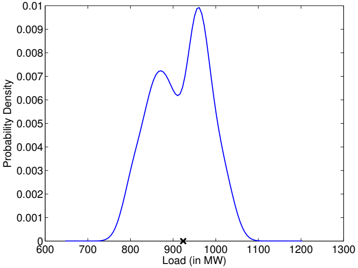


Figure: 10^{th} hour for October 2011

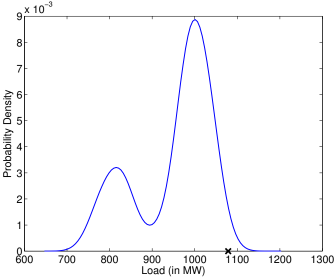


Figure: 18^{th} hour for October 2011

Combination of probability density forecasts

Given N_f probability density estimates i.e., probability density functions $g_{t_f}^{k}(y_{t_f}), \quad k = 1, 2, \ldots N_f$, for at time t_f, of an unknown probability density function, $f_{t_f}(y_{t_f})$

$$\hat{f}_{t_f}(y_{t_f}) = \sum w_{t_f}^{k} \times g_{t_f}^{k}(y_{t_f}) \tag{6}$$

$$\sum_{k=1}^{N_f} w_{t_f}^{k} = 1 \tag{7}$$

$$w_{t_f}^{k} \geq 0 \ \ \forall \ k = 1, \cdots, N_f. \tag{8}$$

Proposed Combination Approaches

- Combining with equal weights
- Combining with Minimization of Weighted Least Squared Error (WLSE)
- Combining with Minimization of Kullback-Leibler Information Criterion (KLIC)

Yogesh K. Bichpuriya, S. A. Soman, and A. Subramanyam, “On optimal combination of probability density forecasts using Kullback-Leibler divergence criterion,” International Conference on Robust Statistics, Halle, Germany, August 2014.

Case Study - Combination of Probability Density Forecasts of Yearly Peak Load

Out-of-sample ex-ante combination of density forecasts of peak load in year 2011-12

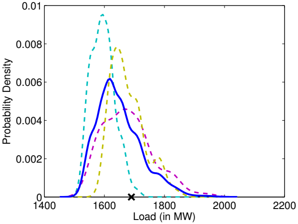


Figure: Equal weights

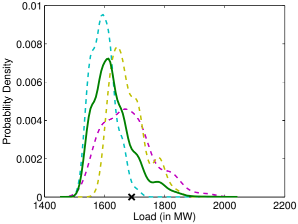


Figure: min WLSE weights

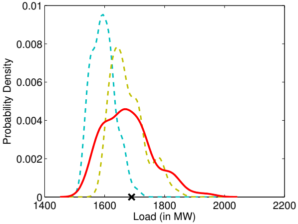


Figure: min KLIC weights

Combination of Point Forecasts

Definition

Let $\mathbf{F}_T = \left[f_{1,T},\ f_{2,T},\ \cdots,\ f_{N_f,T} \right]$ *be a set of* N_f *load forecasts at time* T *by different methods or from different information sets. Then the linear combination of forecasts can be represented as*

$$f_{c,T} = \sum_{i=1}^{N_f} w_{i,T} f_{i,T} = \mathbf{F}_T \mathbf{w}_T \tag{9}$$

where, $\mathbf{w}_T = [w_{1,T}, w_{2,T}, \cdots, w_{N_f,T}]'$ *be an* $N_f \times 1$ *column vector of weights assigned to the individual forecasts at time* T*.*

While some of the combination approaches permit unrestricted weights, others put additional conditions on weights to be non-negative and that their sum to be unity i.e.,

$$\mathbf{u}'\mathbf{w}_T = 1 \tag{10}$$

$$\mathbf{w}_T \geq 0 \tag{11}$$

where, $\mathbf{u} = [1, 1, \cdots, 1]'$ be a unity vector of $N_f \times 1$. Both of these constraints put together implies a convex combination of forecasts.

Proposed Combination Methods

- Median of forecasts
- Weights in proportion to probability of success
- Weights calculation using variance minimization

$$\min_{\mathbf{w}} \frac{1}{2}\mathbf{w}'\mathbf{V}\mathbf{w} \tag{12}$$

subject to:

$$\mathbf{u}'\mathbf{w} = 1 \tag{13}$$

$$\mathbf{w} \geq 0 \tag{14}$$

- Weights calculation using eigenvector of covariance matrix
  - It is an approximation to the above variance minimization problem without solving the quadratic programming problem.
  - We use eigenvector of the covariance matrix of forecast errors to derive weights for combination (*Eigenvec*).
  - Though, this choice of weights will introduce non-optimality, the performance of the method is good.

Yogesh Bichpuriya, S.A.Soman, M.S.S.Rao, "Combination Approaches for Short Term Load Forecasting," 9th International Conference on Power and Energy (IPEC 2010), Singapore, Oct. 2010.

Performance of Combination Approaches for a Discom

Table: Comparison of MAPE and SSE of candidate forecasts and their combinations for the Discom for August 2013 to October 2013

Forecasting Methods		MAPE (%)			SSE ($\times 10^4$)		
Category	Model	Aug	Sep	Oct	Aug	Sep	Oct
Candidate forecasts	ANN	3.11	4.10	5.46	5.44	8.75	15.80
	SDA	3.63	4.24	5.30	7.64	8.78	15.79
	ARIMA	3.60	4.30	5.09	6.29	8.47	14.11
Averaging	Median	2.69	3.62	4.73	3.26	5.93	11.23
Outperformance	PoS	2.70	3.64	4.84	3.30	6.10	11.79
Minimization of variance	MinVarC	2.75	3.80	5.26	3.77	6.81	14.70
	Eigenvec	3.13	3.81	5.13	4.79	6.69	13.42

Yogesh K. Bichpuriya, S. A. Soman, A. Subramanyam, "Combining forecasts in short term load forecasting: Empirical analysis and identification of robust forecaster". *(accepted for publication in SADHANA journal, Springer)*

When the going gets tough, the tough get going

- Missing data and quality of data
- Projection of explanatory variables e.g., weather, economic indicators
- Impact of regulatory policies - open access in distribution system
- Significance of point forecasts or scenario based forecasts in long and medium term
- Increasing penetration of distributed energy resources (e.g., roof-top solar PV)
- Increasing number of electrical/electronic appliances, but more energy efficient
- Plug-in hybrid vehicles - moving loads

We sail within a vast sphere, ever drifting in uncertainty, driven from end to end. -*Blaise Pascal*

Additional References

Breiman, L. and Friedman, J. H. (1985).

Estimating optimal transformations for multiple regression and correlation.
Journal of the American Statistical Association, 80(391):580–598.

Hersbach Hans (2000).

Decomposition of the Continuous Ranked Probability Score for Ensemble Prediction Systems.
Weather and Forecasting, 15(5):559–570.
doi: 10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2.

Silverman, B. W. (1986).

Density Estimation for statistics and data analysis.
Monographs on statistics and applier probability. Chapman and Hall.

Veall, M. R. (1987).

Bootstrapping the probability distribution of peak electricity demand.
International Economic Review, 28(1):203–212.

Thanks

soman@ee.iitb.ac.in