Distribution Feeder Loads Classification and Decomposition

G. Nourbakhsh, H. Y Chiu, Y. Mishra, *Member, IEEE* A. Ghosh, *Fellow Member, IEEE*,

***Abstract*—Load in distribution networks is normally measured at the 11kV supply points; little or no information is known about the type of customers and their contributions to the load. This paper proposes statistical methods to decompose an unknown distribution feeder load to its customer load sector/sub-sector profiles. The approach used in this paper should assist electricity suppliers in economic load management, strategic planning and future network reinforcements.**

***Index Terms*— Load modeling, Load forecasting, Load management**

I. INTRODUCTION

Understanding the usage and the load patterns of customers has always been a major interest in the electricity retail market. Since July 1st 2006, when full contestability took place in Queensland, Australia, an electricity supplier's ability to identify the load details of its users became even more critical. With accurate predictions and classifications of the customer load types, the electricity suppliers can better manage unpredicted situations, such as load forecasting and transformer overloads. The network planners and operators will also benefit from the customer load findings in areas such as load diversification, reinforcement, upgrading and load ratings of transformers.

While there are many benefits associated with understanding the customer load types, very few publications seem to be focused in this area. One possible explanation may relate to the costs of acquiring data for analysis. This can be due to the fact that, although customer load types can be obtained by placing measuring devices on the 11KV feeders, but this can be a very expensive exercise for utilities to get the information.

Basu [1] provided some insights on the topic of classifying loads into customer profiles. Basu conducted surveys and from the data obtained separated the load feeder into components of residential, commercial and industrial. Since the classification process was not cited in his paper, therefore a comparison between Basu's method and the methods in this paper was not possible. The next important difference is that Basu focused on load prediction rather than on customer profiles and feeder decomposition.

This paper uses a technique in classifying the feeders into different customer load profiles and then uses this information to decompose unknown feeder loads to find the customer load type ingredients.

II. DATA COLLECTION

Three types of customer loads; residential, commercial and industrial were considered in this paper. The final estimated results can be affected by the initial information collected from the known feeder load types. Accuracy can be obtained by applying the following constrains on the half hourly feeder loads under study. The constrains are given as follows:

- ✓ In the customer profile process the selected known feeder loads need to be composed largely of just one type of customer from one of the three sectors.
- ✓ Within the data collection period the selected known feeder loads should have sufficient data sets to reduce the effect of possible errors; missing load values and abnormalities in the measurements.
- ✓ Within the data collection period the selected known feeder loads should not have gone through any major upgrades or changes.
- ✓ To improve the model's ability to generalize, known feeder loads were selected from a range of different area locations.

While all constrains were taken into consideration, 34 feeder loads in total (11 from each sector and 1 unknown feeder for section V; decomposition) selected for the study.

III. MODEL FORMULATION

The half-hourly load profile of a feeder describes how one unit of ampere (directly relating to power) is drawn throughout a week. The load profile of a given feeder is expressed as a sequence of proportions:

$$p_{t,d}, \quad t = 1,\ldots,48, \quad d = 1,\ldots,7,$$

where t denotes the half-hourly time interval and d denotes the day of the week (Monday to Sunday). The element $p_{t,d}$ of a load profile gives the proportion of the weekly load that is drawn during the t-th half-hourly interval of day d. We express this property mathematically as in (1):

$$\sum_{d}\sum_{t} p_{t,d} = 1, \quad p_{t,d} \geq 0. \quad (1)$$

In order to reduce the number of parameters that need to be estimated in a load profile, it is assumed that the load drawn for each weekday follows the same pattern. This is expressed mathematically as

$$p_{t,1} = \cdots = p_{t,5}$$

for all t. The loads drawn on Saturday and Sunday are permitted to differ from each other and from the load drawn on the weekdays. These basic assumptions appear consistent with graphical inspection of the load data (see Fig. 1 and Fig. 2). No further assumptions are made concerning the structure or functional form of the load profile.

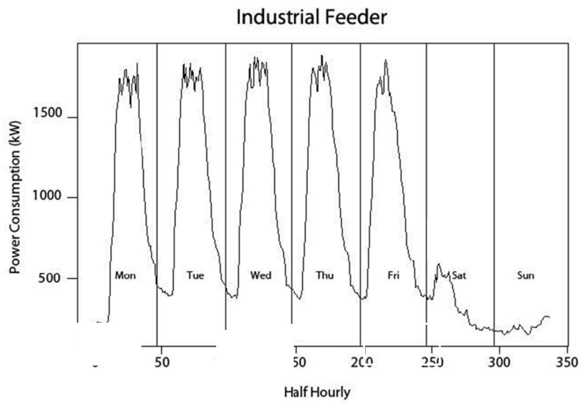

Fig. 1 Industrial feeder plot for 1 week

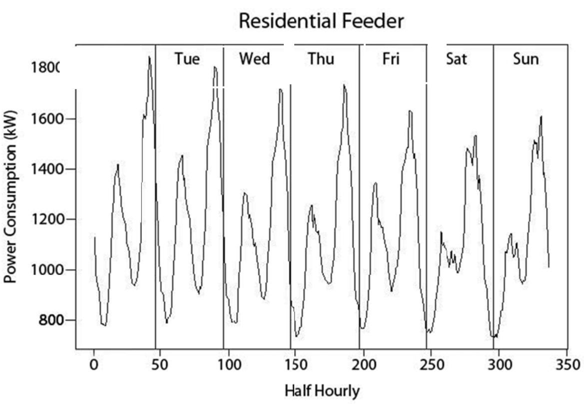

Fig. 2 Residential feeder plot for 1 week

The load profile only describes the proportion of weekly load drawn during each of the half-hour intervals. In order to relate the load profile to the data we need to introduce a new variable m_w which models the amount of current drawn during week w and can be written as (2). The observed amount of current drawn on the feeder at time t on day d and week w is modeled by (3).

$$m_w \geq 0, \quad (2)$$

$$y_{t,d,w} = m_w p_{t,d} + e_{t,d,w}, \quad (3)$$

where $e_{t,d,w}$ denotes the stochastic variation in the current drawn that is not explained by the weekly expected load and load profile. This stochastic component is assumed to be white noise with mean zero and constant, finite variance. We consider this model to be as simple as possible while capturing the important characteristics of the load feeder. The most important assumption for this model is that the load profile does not change during the data collection period. Therefore any change in the load pattern must be explained by the expected weekly load m_w. For example, if growth occurs in an area supplied by a feeder then the additional load is assumed to have a similar pattern to the original.

The parameters of this model, the load profile $p_{t,d}$ and weekly expected load m_w, are estimated using least squares. Due to the constraints (1) and (2), the standard Gauss-Newton algorithm cannot be applied. Fortunately, the bi-linear form of the model (3) leads to a particularly simple algorithm for finding the least squares estimate.

Initialize m: For all w set

$$m_w = 7 \times 48 \times \bar{y},$$

where $\bar{y}$ is the sample mean of the feeder load data.

Repeat until convergence:

1. *Update profile*: Solve

$$\min_{p_{t,d}} \sum_{t,d,w} \left(y_{t,d,w} - m_w p_{t,d} \right)^2, \quad (4)$$

subjected to the constraints (1).

2. *Update weekly expected load*: Solve

$$\min_{m_w} \sum_{t,d,w} \left(y_{t,d,w} - m_w p_{t,d} \right)^2, \quad (5)$$

subjected to the constraints (2).

In the *Update profile* step of the algorithm, given an estimate of the weekly expected load, the profile can be estimated solving a linear least squares problem subject to a linear equality constraint and non-negative constraints. Algorithms for solving this problem can be found in [3, 4, 5]. In the data analyzed, the non-negativity constraint did not need to be explicitly enforced and so the profile estimate could be obtained by solving the standard linear least squares problem.

In the *Update weekly expected load* step of the algorithm, given an estimate of the profile the weekly expected load can be estimated by solving a linear least squares problem subject to non-negativity constraints. As the weekly expect load parameters are orthogonal, the solution to this problem is simply

$$m_w = \begin{cases} 0, & \hat{m}_w < 0, \\ \hat{m}_w, & \hat{m}_w \geq 0, \end{cases}$$

where $\hat{m}$ is the solution to (5) without the constraints (2). As was the case when updating the profile, the non-negativity constraints never had to be explicitly enforced for the data analyzed in this paper. The two steps of the algorithm are iterated until convergence is reached. Convergence is assessed by changes in residual squared error and by changes in the parameter estimates.

Using this algorithm, the load profile and expected weekly load were estimated. The predicted load values generated from the model and the actual load values from the feeder load are plotted in Fig. 3. The model appears to fit the feeder load data quite well, except for some large anomalous peaks. The anomalous peaks might be due to sudden changes in temperature from one day to the next, the breakdown of the feeder supplying to the neighboring area causing the feeder under study abnormal overloaded, or a ceremonial/public holiday. Peaks of this type are not modeled by (3). However, provided that such events are rare, the effect of these peaks on the estimated profile should be minimal.

The profile for this feeder is shown in Fig. 4. This process was repeated for each of the 33 load feeders of homogeneous sectors. The variation in the profiles of the 11 industrial load feeders is shown in Fig. 5.

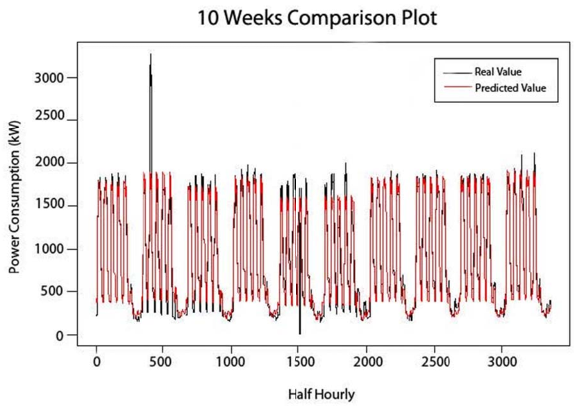

Fig. 3 Predicted results versus Actual results

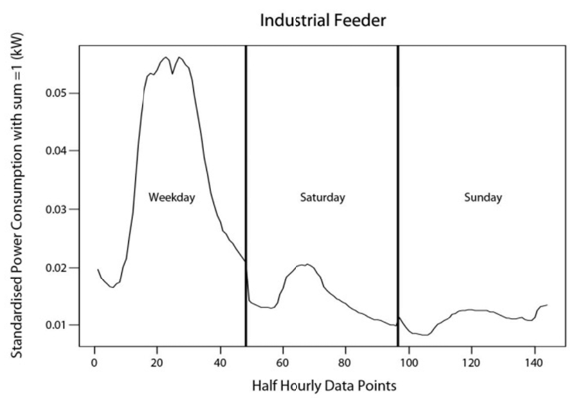

Fig. 4 Weekday, Saturday and Sunday load plot for a single industrial feeder.

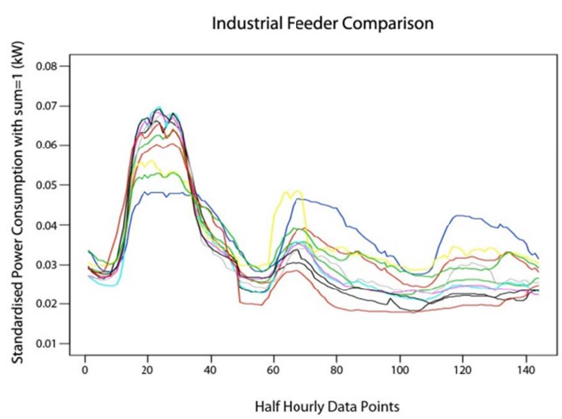

Fig. 5 Industrial loads Comparison

Understanding the uncertainty in parameter estimation is an important part of any statistical analysis. For the model proposed in this paper, one could possibly construct confidence intervals and error bands on the feeder profiles using some form of bootstrap [7] or using the limiting distribution of the estimates. However, this is not a primary concern for the current paper since the feeder profiles are not of direct interest and are only used as an intermediate step in the decomposition of the feeders. Of greater importance to the current work is the accuracy of the decomposition method which will be investigated using cross-validation in section VI.

IV. SUB-SECTOR /CUSTOMER PROFILES

So far we have described how load profiles for given feeders can be estimated from observed load data. We now aim to determine load profiles for a given sector or sub-sector. Each feeder load profile is affected to some extent by local factors that are of little importance to the general sector to which it belongs. To obtain sector profiles we average over the feeder load profiles in each sector. This averaging reduces the effect of variation in feeder load profiles on the sector profile. It is easily verified that the sector profiles satisfy the conditions (1) imposed on load profiles.

One problem that may arise in the above procedure is that the classification of the areas that the feeders serve into industrial, residential and commercial is too coarse. In that case, the average of the feeder load profiles is not representative of the load profiles in the sector. This problem can be remedied by performing some type of clustering of the feeder load profiles within each sector to identify appropriate sub-sectors. In this paper we use the K means clustering algorithm where the distance between two feeder loads is measured by the Euclidean metric. Various criteria have been proposed to determine the most appropriate number of clusters, for example [8]. However, the aim of this paper is to determine percentages associated with each sector, the number of clusters by the performance in prediction as assessed by cross-validation are determined.

Fig. 6 illustrates three sub-sector load profiles obtained from the commercial sector. This information along with the sub-sector load profiles established from the industrial and residential sectors forms the bases used to decompose an unknown feeder load.

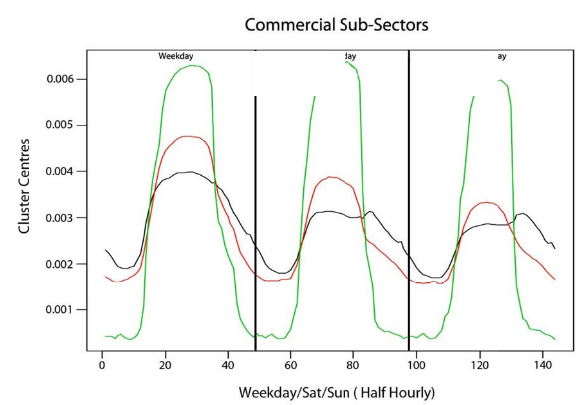

Fig. 6 Sub-sectors within commercial

In this section we have introduced an assumption that is crucial to our analysis: the load pattern of any feeder follows the same basic behaviour of a sector or sub-sector. If a single feeder load operates in completely different fashion to all other feeders, then it will not be possible to form accurate sector or sub-sector profiles. This situation may arise, for example, if a load feeder supplies to an industrial company that operates 24 hours 7 days a week whereas all other observed industrial load feeds supply industries where operation hours are from 7:30 to around 18:00, 5 days a week.

V. DECOMPOSING A LOAD FEEDER

In the previous section, load profiles were identified for each of the observed subsectors. These load profiles were built on the assumption that the area serviced by each feeder was homogeneous, that is, only one sector was predominant in the area serviced. However, feeders will often serve an area that comprises a number of different subsectors. Therefore, a method is needed which is able to determine which subsectors are present in the area serviced by the feeder and how much each subsector contributes to the load drawn.

Assume that, using the algorithm described in Section IV, a characteristic load profile for each subsector is identified. The load profiles of the subsectors is denoted by

$$P^1_{t,d}, \ldots, P^S_{t,d},$$

where S is the number of subsectors. It is assumed that the load profile of the feeder load for the new area can be represented by (6) subjected to the constraint (7)

$$p_{t,d} = \omega_1 P^1_{t,d} + \cdots + \omega_m P^m_{t,d}, \quad (6)$$

$$\omega_1 + \cdots \omega_m = 1, \quad \omega_i \geq 0. \quad (7)$$

The coefficients ω_i give the proportion of the load that each subsector draws. In particular, if ω_i is zero, then the i-th subsector does not make any contribution to the load drawn by this feeder. The coefficients ω_i in equation (6) are estimated by least squares subject to the constraints (7). Hence, they can be estimated using the algorithm described in [3, 4, 5].

As an example, this method is applied to a feeder servicing an area whose composition in terms of the subsectors is unknown. The load profile of the feeder is first determined using the method described in Section III. The load profile is then decomposed into the subsector load profiles. The estimated coefficients ω are given in Table I.

Sub-Sectors/Profiles	Correlation Values
Ind1	0.1433486
Com1	0.1489686
Com2	0.02814233
Com3	0
Res1	0.6795404

Table I. Contribution of subsector Profiles

From Table I, it can be seen that this feeder supplies mostly to the residential sector. The estimated coefficient for then subsector Com3 is zero. This implies that the Com3 subsector is not present in the area supplied by the feeder.

Having determined the relative contributions of each subsector load profile, it is possible to plot the contribution of each subsector to the average weekly load. This is achieved by multiplying the respective load profiles by the product of their relative contribution ω and the average of the weekly multiplier for the feeder load under study. Fig. 7 shows the subsector contributions of the load feeder under study. It is worth noting that the sum of load contributions closely follows the real load values with an overall difference of less than 1%.

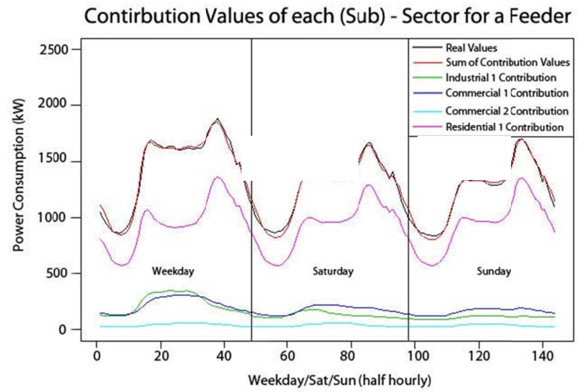

Fig. 7 Contribution values of Sub-Sectors for an unknown feeder

VI. CROSS VALIDATION AND TESTING

Cross validation is a technique by which an approximately unbiased estimate of the predictive accuracy of an estimated model is obtained. The method has a long history in statistics [9] and widely accepted in the statistics and machine learning communities. The technique involves splitting the data set into two groups, a training set and a test set. The parameters of the model are estimated using the data in the training set. The resulting model is then used to predict the test set and the accuracy of these predictions measured. This can be repeated over different partitions of the data in training and test sets with the results average to provide an approximately unbiased estimate of accuracy. In this paper we shall use leave-out-one cross validation. While leave-out-one cross validation generally refers to removing a single observation from the data set, in this analysis we will be removing one load feeder at a time to form the training set. The test set is that load feeder that was removed. This procedure is repeated for each of the 33 homogeneous load feeders that were observed.

Firstly, the case where the sector load profiles are formed by taking averages of the feeder load profiles is considered. Using the method of Section V, the load feeder is decomposed into the three sectors. As each load feeder in the data set is believed to be homogeneous, the estimated proportions ω_i should be close to one for one of the sectors and close to zero for the other two. For some load feeders this is not observed. This may be due to variation in the load feeder which is not reflected by the sector load profile averages. In any case, as we are dealing with load feeders that are believed to be homogeneous, a load feeder is classified according to the largest estimated proportion ω_i. The following misclassifications are observed: 1 for Industrial, 4 for Commercial and 0 for Residential. Cross Validation Results for Industrial, commercial and residential loads are shown in Fig.8, Fig. 9 and Fig.10 respectively.

As the commercial sector appeared to be most problematic, the possibility of subsectors within the commercial sector is considered. Two and three possible subsectors were considered and their accuracy assessed by cross-validation. A load feeder is now classified according to the sector with the largest estimated proportion, summing over its subsectors. When the procedure is applied with two subsectors for the commercial sector, the number of misclassifications was increased: 3 for Industrial, 6 for Commercial and 0 for Residential. However, applying the procedure with three subsectors for the commercial sector, the fewer misclassifications are observed: 3 for Industrial, 0 for Commercial and 0 for Residential.

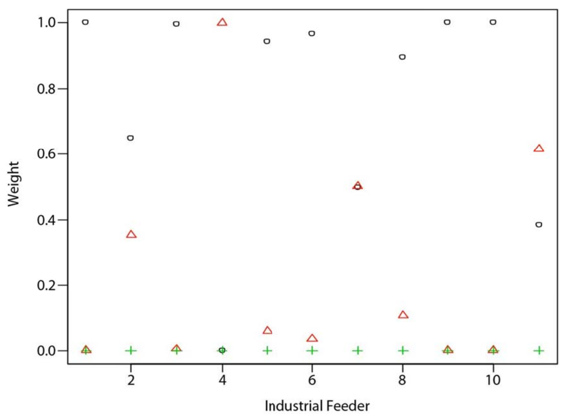

Fig. 8 Cross Validation Results for Industrial

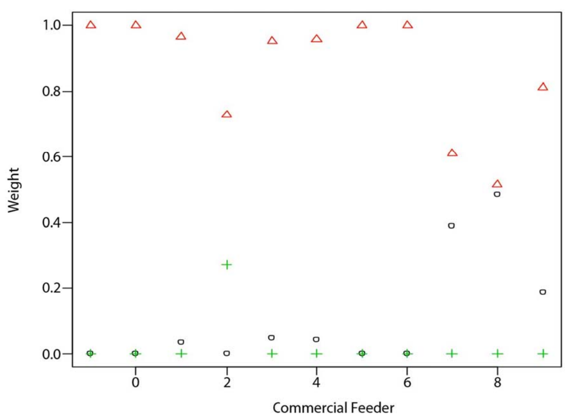

Fig. 9 Cross Validation Results for Commercial

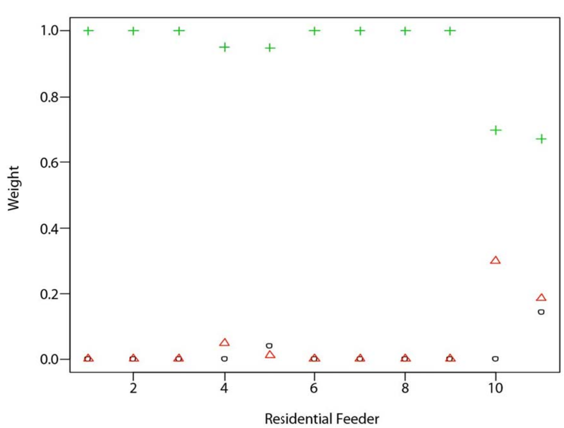

Fig. 10 Cross Validation Results for Residential

The results from the classifications with three subsectors for the commercial sector are summarized in the plots below. In these plots a circle represents the weight assigned to the Industrial sector, a triangle the weight assigned to the Commercial sector and a cross the weight assigned to the Residential sector. Ideally, all weights would be concentrated near one or zero as appropriate.

Further investigation into this issue discovered that the reason for the commercial misclassifications may be caused by the similar load pattern of the commercial and residential sectors. To cater for the shopping needs of the residential customers many commercial shops and businesses have altered their opening hours to suit those of the residential sector. One example is the grocery chain Woolworth that has changed their opening hours to include Saturday and Sunday for the convenience of the customers.

The other misclassifications where an industrial feeder was mistaken as a commercial feeder and a commercial feeder as an industrial feeder can be explained after the feeder loads were compared with its respective sector feeder loads. After comparison it was found that those feeder loads demonstrated completely different load patterns to the other feeder loads in its respective sector. Since the feeder loads were unique and different from all the other feeder loads in its sector, therefore the model can not accurately identify the type of sector the feeder load belonged to. With this discovery it is possible to classify these feeder loads as unique sector feeders for future study references.

Other measures of accuracy besides the 0 – 1 loss could be considered. Of particular interest are measures that are more sensitive to the error in the estimated proportion. This aspect was not investigated since we only have the classification of the load feeders when they are homogeneous.

VII. Conclusions and Recommendations

This paper examined an approach that would allow electricity suppliers to distinguish the type of (sub)-sector/customer load profiles contributing to the 11KV loads without installing measuring devices that can be very expensive.

Although this simple model has some imperfections when classifying subsector/customer load profiles, the methodology has considerable potential. One area requiring further investigation is the incorporation seasonal, weather and temperature variations into the load profiles. Another area of investigation is the accuracy assessment of the estimated proportions in the decomposition (6). This will require a combination of additional data and expert knowledge against which the estimates can be verified.

With up-to-date and adequate information on subsector/customer load profiles and using the method described in this paper, it is possible to find the types of load and their percentages contributed to an unknown feeder.

This application can provide multiple benefits for electricity suppliers in areas such as; economic load management such as load diversification relating to electricity block purchasing, load forecasting, retailing and strategic network planning.

VIII. References

[1] S. N. Basu, "Short Term Localized Load Prediction", IEEE Transactions on Power Systems, Vol. 7, No. 1, pp. 389-397, 1992.
[2] G.A.F. Seber and C. J. Wild, (1988) *Nonlinear Regression,* New York: Wiley.

[3] R. Bro and S. De Jong 'A fast non-negatively-constrained least squares algorithm' Journal of Chemometrics, 11, pp. 393-401, 1997

[4] C. L. Lawson and B. J. Hanson, (1974), *Solving Least Squares Problems*, New York: Prentice-Hall

[5] S. N. Wood, "Monotonic smoothing splines fitted by cross validation", SIAM Journal on Scientific Computing, Vol. 15(5), pp.1126-1133, 1994

[6] T. Hastie, R. Tibshirani and J. Friedman, (2001) Elements of Statistical Learning: data mining, inference, and prediction, New York: Springer.

[7] D.N. Politis, 'The impact of bootstrap methods on time series analysis,' Statistical Science, 18, pp. 219-230, 2003

[8] R. Tibshirani, G. Walther and T. Hastie, "Estimating the Number of Clusters in a Data Set via the Gap Statistic", Journal of the Royal Statistical Society, Series B (Statistical Methodology), Vol. 63, pp. 411-423, 2001

[9] S. Arlot and A. Celisse 'A survey of cross-validation procedures for model selection', Statistics Surveys, pp. 40-79, 2010