



Yüksel Öner

Ondokuz Mayıs University, yoner@omu.edu.tr, Samsun-Turkey

Fikriye Kabakçı

Recep Tayyip Erdoğan University, fikriye.kabakci@erdogan.edu.tr,
Rize-Turkey

Burçin Öner

Mehmet Gürcan

Fırat University, mgurcan@firat.edu.tr, Elazığ-Turkey

DOI	http://dx.doi.org/10.12739/NWSA.2019.14.4.2A0176	
ORCID ID	0000-0003-2433-3304	0000-0001-6266-1902
	0000-0001-9550-0435	0000-0002-3641-8113
CORRESPONDING AUTHOR	Yüksel Öner	

A REMARK ON L2 DISTANCE FUNCTION AND NON-IDENTIFIABILITY PROBLEM OF FINITE MIXTURE DISTRIBUTION MODELS IN MODEL-BASED CLASSIFICATION

ABSTRACT

Finite mixture models provide flexible method of modeling data obtained from population consisting of finite number of homogeneous subpopulations. One of the main areas in which the finite mixture model structures is practically used in statistics is model based classification. However, the result of non identifiability problem arising from the structure of the finite mixture models may cause unreliable results on classification. In this paper we compare the probability density functions (*pdfs*) of the finite mixture distribution models for two different populations by L2 distance. We propose the componentwise L2 distance function to compare the *pdfs* of finite mixture distribution models for two different populations in the presence of non identifiability problem. Besides, a condition is proposed to control whether the L2 distance function gives similar results with the componentwise L2 distance function to compare the *pdfs* of finite mixture distribution models for two different populations.

Keywords: Finite Mixture Distribution, L2 Distance Function, Model Based Classification, Mixture Model, Non-identifiability

1. INTRODUCTION

Finite mixture models have continued to receive increasing attention recently. Major application areas of finite mixture models are economics, financial, medicine, biology, genetics and social sciences. In addition of its importance in applications, finite mixture models constitute an efficient and challenging area of statistical research. It has widely applied to classification and clustering. Classification is the process of assigning group membership labels to unlabeled observation [1]. McNicholas has pointed out that there are three types of classifications based on the supervision level; supervised, semi-supervised, and unsupervised [1]. Unsupervised classification is known as clustering and no observations are a priori labelled, and the other two species have some observations that are priori labeled to use this knowledge to infer labels for the unlabelled observations. Semi-supervised classification process which is described by the use of a statistical model is termed "model-based classification" in the literature. One of the application

How to Cite:

Öner, Y., Kabakçı, F., Öner, B., and Gürcan, N., (2019). A Remark on L2 Distance Function and Non-Identifiability Problem of Finite Mixture Distribution Models in Model-Based Classification, Technological Applied Sciences (NWSATAS), 14(4):139-146, DOI:10.12739/NWSA.2019.14.4.2A0176.



areas of the model-based classification is the comparison of two different populations which formed by the finite mixture models.

The probability density functions of the mixture distribution models for two different populations can be compared by using distance functions between them in model-based classification applications. In many cases the component densities of the finite mixture distribution models for two different populations are in irregular order and they act as permutation functions. This problem is known as non-identifiability problem of finite mixture distribution models. The problem of non-identifiability is a general term that needs to be configured depending on the context. The identifiability of a parameter is explained by the fact that the corresponding parameters are equal if the any two members of the family of distributions are equal. However, this may not always be the case. When one-dimensional normally distributed model is considered, it is known that variances with opposite signs produce the similar normal probability density functions which exemplifies the non-identifiability of parameter. In other words, estimation of parameters based on observations meaningful when parameter is identifiable [2]. In addition, the definition of identifiability for finite mixture models slightly different and explained in the following section. Erol suggested to take into account non-identifiability problem of finite mixture distribution models [3]. For this, he proposed to componentwise distance function to compare the *pdfs* of the finite mixture distribution models for two different populations if the component densities are permutation functions. In the case of component densities are in irregular order and there is a problem in the comparison of *pdfs* of the finite mixture distribution models for two different populations. Erol suggested to compare by using component Hellinger distance function between them [3 and 4].

2. RESEARCH SIGNIFICANCE

We suggest to use L_2 distance and component wise L_2 distance to compare *pdfs* of the finite mixture distribution models for two different populations. We give the definition of identifiability problem of mixture models in detail and the condition theorem by using L_2 distance with the proof in chapter 3, we summarize our results in chapter 4.

3. MATERIAL AND METHOD

The formal definition of finite mixture model which is mentioned in the introduction part is given by Titterington et al. as follows [5].

- **Definition 1:** The distribution of a random variable X with density function of the form

$$f(x, \theta) = \sum_{j=1}^k w_j f(x, \theta_j) \quad (1)$$

is named a finite mixture distribution with k components and $\theta = (w_1, \dots, w_k, \theta_1, \dots, \theta_k)$ parameter vector. Here $f(x, \theta_j)$, $j = 1, \dots, k$ denote the component densities of mixture with parameter θ_j and mixing weights w_1, \dots, w_k are positive and sum to unity. Component densities $f(x, \theta_j)$, $j = 1, \dots, k$ are not necessary to come from similar distributions. However, we assume that component densities are Gaussian.

- **Definition 2:** A random variable X with density function

$$f(x, \theta) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

where $\theta = (\mu, \sigma^2)$, is said to be normal or Gaussian distributed



with parameters μ and σ^2 . Recommended common methods for estimating model parameters in a finite mixture model is to apply Expectation Maximization algorithm [6]. The first and proper theoretical study of the EM algorithm proposed by Dempster et al., [6]. EM algorithm is a method based on iteration, which provides numerical approach to the maximum likelihood estimation of the parameters in mixture models with the help of hidden variables. Studies shown that the EM algorithm is not only slow in convergence but also unstable against outliers [7]. To overcome the limitations of EM algorithm, minimum distance estimation method based on integrated squared error criterion, termed L2E, has been proposed as a more robust method compared to the EM algorithm by Scott [8]. In addition to parameter estimation of finite mixture models, a study on finding the number of mixture components proposed by Thayasivam [9]. Fitting a mixture distribution to the data from given observations x_1, \dots, x_n , requires identifiable parameter vector θ . This is because unique characterization exists for any mixture model when the parameter vector θ is identifiable.

In general, a parametric family of density functions $f(x, \theta)$ is identifiable if distinct values of the parameter θ determine distinct members of the family. That is,

$$f(x, \theta) = f(x, \theta^*) \tag{2}$$

if and only if $\theta = \theta^*$ [2]. Identifiability definition of finite mixture models is given slightly different. To see why this is necessary, suppose that $f(x, \theta)$ has component densities $f(x, \theta_j)$ and $f(x, \theta_j^*)$, that belong to the same parametric family. Then the equation (2) will still hold when the component labels j and j^* are interchanged in θ . Although this class of mixtures may be identifiable, θ is not. Hence, identifiability of mixtures can be explained as follows.

- **Definition 3:** Let $f(x, \theta) = \sum_{j=1}^k w_j f(x, \theta_j)$ and $f(x, \theta^*) = \sum_{j=1}^{k^*} w_j^* f(x, \theta_j^*)$ be any two members of a parametric family of mixture densities. This class of finite mixtures is said to be identifiable for θ if $f(x, \theta) = f(x, \theta^*)$ if and only if $k = k^*$ and the component labels can be permuted so that $w_j = w_j^*$ and $f(x, \theta_j) = f(x, \theta_j^*)$. Interchanging of component labels cause non-identifiability problem of θ . It is proposed to add appropriate constraints on θ to solve this problem. Aitkin and Rubin suggested to sort the component labels from smallest to largest as $w_1 \leq w_2 \leq \dots \leq w_k$ and ignored this order in the parameter estimation process by MLE [10]. McLachlan and Peel suggested to consider similar idea on the component labels when the component distributions are univariate normal [2]. Some other detailed definitions and recommendations about identifiability concept are given by Teicher, Yakowitz and Spragins, Kadane, Akdağ, Toher et al., McLachlan and Basford, Fraley et al., Erol and Akdeniz, Dean et al., Wehrens et al., Yeung et al., Servi, Titterington et al. [11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22 and 5]. Non-identifiability problem needs to be considered when comparing the populations represented by finite mixture models. It has been suggested in the literature that *pdfs* of the finite mixture distribution models are arranged in ascending order according to their means. However, such sorting can take place differently in different populations. Erol explained this situation by the popular

example which includes two different populations each consisting of boys and girls and assuming each population has the mixture of two univariate normal distributions w.r.t the variate height in centimeter [3 and 23]. Systolic Blood Pressure (The highest pressure when the heart beats and pushes the blood round the body) of female and male patients from different populations could be different example for the similar purpose of comparison. As Erol pointed out there are two type of scenarios we encounter when we compare the pdfs of the finite mixture models for two different populations by distance function [3]. We generated populations from normally distributed two-component finite mixture models to illustrate non-identifiability problem.

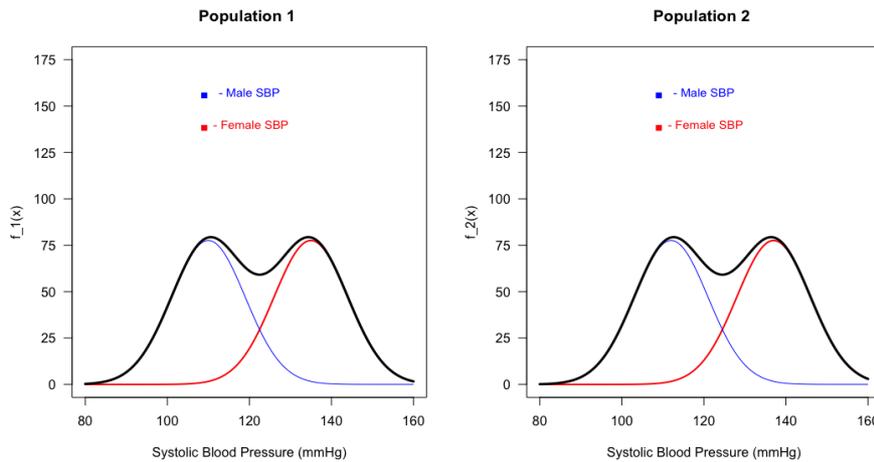


Figure 1. Finite mixture distribution models and their components for two populations

Corresponding component densities of the mixture distribution models for two different populations are in the same order with respect to their means, as shown in Figure 1. Therefore, there is no need to draw attention to the non-identifiability problem. However, for the second case non-identifiability problem is clearly observed in the Figure 2.

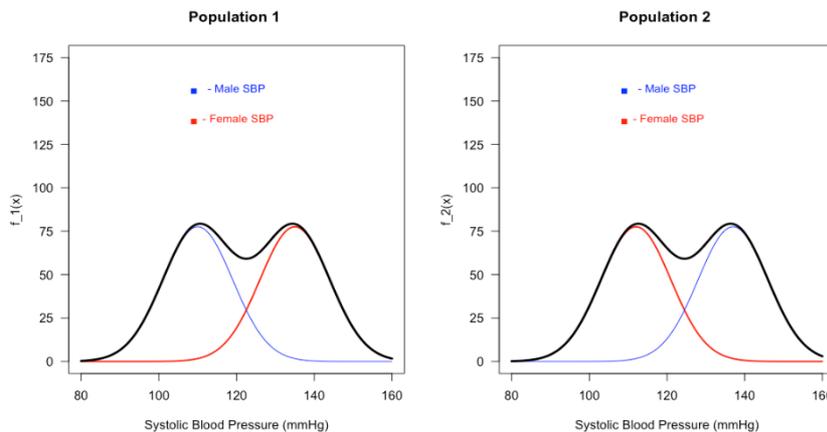


Figure 2. Finite mixture distribution models and their components for two populations

Component densities of mixture distribution models are not in the same order with respect to their means for population 1 and



population 2, and they act as permutation functions, as shown in Figure 2. Erol suggested to componentwise Hellinger distance function to compare the populations with irregular component densities. Besides, he suggested a condition under which the value of Hellinger distance function between the *pdfs* of the finite mixture distribution models is equal to the value of componentwise Hellinger distance function between *pdfs* of the finite mixture distribution models for two different populations [3].

4. FINDINGS AND DISCUSSIONS

In our study we will restrict our attention to the L_2 distance between *pdfs* of the finite mixture distribution models for two different populations. The details of mixture models for two populations from univariate normal distribution with two components and L_2 distance will be given first, then the definitions and conditions defined by Hellinger distance will be adapted to the L_2 distance. Let the mixture model for the i -th population with the mixture of two univariate normal distributions is given as follows;

$$f_i(x; \mu_{i1}, \mu_{i2}, \sigma_{i1}, \sigma_{i2}, p_{i1}) = \sum_{j=1}^2 w_{ij} f_{ij}(x; \mu_{ij}, \sigma_{ij})$$

$$-\infty < x < +\infty, \quad -\infty < \mu_{ij} < +\infty, \quad \sigma_{ij} > 0 \quad (3)$$

Here i refers to population number and j refers to each component of mixtures. If the component densities are normally distributed densities become

$$f_{ij}(x; \mu_{ij}, \sigma_{ij}) = \frac{1}{\sqrt{2\pi}\sigma_{ij}} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu_{ij}}{\sigma_{ij}}\right)^2\right\}$$

$$-\infty < x < +\infty, \quad -\infty < \mu_{ij} < +\infty, \quad \sigma_{ij} > 0 \quad (4)$$

Where μ_{ij} and σ_{ij}^2 are the j -th component mean and the variance of the i -th population respectively. w_{ij} 's are mixture proportions such that

$$\sum_{j=1}^2 w_{ij} = 1 \text{ for } i = 1, 2 \text{ [3].}$$

If the corresponding *pdfs* of finite mixture distribution models for two populations can be sorted based on their means such that;

$$\mu_{11} < \mu_{12} \text{ and } \mu_{21} < \mu_{22} \quad (1)$$

There is no problem for comparing *pdfs* of finite mixture distribution models for two different populations. By keeping the same notations with the Erol's paper the *pdfs* of finite mixture distribution models for two populations $f_1(x; \mu_{11}, \mu_{12}, \sigma_{11}, \sigma_{12}, p_{11})$ ($f_1(x)$ for short) and $f_2(x; \mu_{11}, \mu_{12}, \sigma_{11}, \sigma_{12}, p_{11})$ ($f_2(x)$ for short) can be compared by L_2 distance between them. L_2 distance between $f_1(x)$ and $f_2(x)$ is given as follows [3];

$$L_{2f_1, f_2} = \int_R (f_1(x) - f_2(x))^2 dx \quad (2)$$

If the corresponding *pdfs* of finite mixture distribution models for two populations are not in the same order w.r.t their means such that

$$\mu_{11} < \mu_{12} \text{ and } \mu_{21} > \mu_{22} \text{ or } \mu_{11} > \mu_{12} \text{ and } \mu_{21} < \mu_{22} \quad (3)$$

Then the component densities are in irregular order and the corresponding *pdfs* of finite mixture distribution models act as permutation functions. In the case of component densities are in irregular order and there is a problem in the comparison of *pdfs* of the finite mixture distribution models for two different populations Erol suggested to compare by using componentwise Hellinger distance function between them [3]. We suggest to use L_2 distance and

componentwise L_2 distance to compare *pdfs* of the finite mixture distribution models for two different populations.

$$CL_{2(f_1, f_2)} = \sum_{j=1}^2 \int_R (w_{1j}f_{1j}(x) - w_{2j}f_{2j}(x))^2 dx \quad (4)$$

$$R = \{x | -\infty < x < +\infty\}$$

Considering the condition given by Hellinger's distance function, we give following theorem that shows the L_2 distance to compare *pdfs* of the finite mixture distribution models and component wise L_2 distance to compare *pdfs* of the finite mixture distribution models for two different populations can be equal.

Theorem 4.1. L_2 distance function value of the *pdfs* of the finite mixture distribution models for two different populations is equal to the component wise L_2 distance function value of the *pdfs* of the finite mixture distribution models if and only if the equality

$$\int_R (w_{11}f_{11} - w_{21}f_{21})(w_{22}f_{22} - w_{12}f_{12}) = 0 \quad (9)$$

holds.

Proof: We give the proof only for if part. L_2 distance between the *pdfs* of the finite mixture distribution models $f_1(x)$ and $f_2(x)$ given by

$$L_{2(f_1, f_2)} = \int_R (f_1(x) - f_2(x))^2 dx \quad (10)$$

the equation (10) can be written as

$$L_{2(f_1, f_2)} = \int_R f_1(x)^2 - 2f_1(x)f_2(x) + f_2(x)^2 \quad (11)$$

In terms of component densities as

$$L_{2(f_1, f_2)} = \int_R (w_{11}f_{11}(x) + w_{12}f_{12}(x))^2 dx \quad (12)$$

$$+ \int_R (w_{21}f_{21}(x) + w_{22}f_{22}(x))^2 dx - 2 \int_R f_1(x)f_2(x) dx$$

We add and subtract the terms $2w_{11}f_{11}w_{21}f_{21}$ and $2w_{12}f_{12}w_{22}f_{22}$ in the integral to the equation (12) and we obtain the following equation

$$L_{2(f_1, f_2)} = \int_R ((w_{11}f_{11})^2 + (w_{21}f_{21})^2) dx \quad (13)$$

$$- 2 \int_R (w_{11}f_{11}w_{21}f_{21}) dx + \int_R ((w_{12}f_{12})^2 + (w_{22}f_{22})^2) dx$$

$$- 2 \int_R (w_{12}f_{12}w_{22}f_{22}) dx + 2 \int_R (w_{11}f_{11}w_{21}f_{21}) + 2 \int_R (w_{12}f_{12}w_{22}f_{22}) dx$$

$$+ 2 \int_R (w_{11}f_{11}w_{12}f_{12}) dx + 2 \int_R (w_{21}f_{21}w_{22}f_{22}) dx - 2 \int_R f_1(x)f_2(x) dx$$

The first two terms in equation (13) are equal to the componentwise L_2 distance function between the component densities of the *pdfs* of the finite mixture distribution models for two different populations. Therefore, we have

$$L_{2(f_1, f_2)} = \sum_{j=1}^2 \int_R (w_{1j}f_{1j} - w_{2j}f_{2j})^2 dx \quad (14)$$

$$+ 2 \int_R (w_{11}f_{11}w_{21}f_{21}) + 2 \int_R (w_{12}f_{12}w_{22}f_{22}) dx$$



$$+2 \int_R (w_{11}f_{11}w_{12}f_{12})dx + 2 \int_R (w_{21}f_{21}w_{22}f_{22})dx - 2 \int_R f_1(x)f_2(x)dx$$

or

$$L_{2(f_1, f_2)} = CL_{2(f_1, f_2)} + 2 \int_R (w_{11}f_{11}p_{21}f_{21}) + 2 \int_R (w_{12}f_{12}w_{22}f_{22})dx \tag{15}$$

$$+2 \int_R (w_{11}f_{11}w_{12}f_{12})dx + 2 \int_R (w_{21}f_{21}w_{22}f_{22})dx - 2 \int_R f_1(x)f_2(x)dx$$

By expanding the last term in equation (15) we obtain the following equation

$$L_{2(f_1, f_2)} = CL_{2(f_1, f_2)} - 2 \int_R (w_{11}f_{11} - w_{21}f_{21})(w_{22}f_{22} - w_{12}f_{12}) \tag{16}$$

The value of L_2 distance function between the *pdfs* of the finite mixture distribution models is equal to the value of componentwise L_2 distance function between of the *pdfs* of the finite mixture distribution models for two different populations if the condition in equation (9) holds.

5. CONCLUSIONS AND RECOMMENDATIONS

In this article, we noted that *pdfs* of the finite mixture distribution models for two different populations may have irregular order. This is known as non-identifiability problem of finite mixture distribution models. In this case, comparing these two populations may not yield reliable results. Therefore, this problem must be considered in the application areas of finite mixture models such as classification and clustering. The theorem we proposed here is a mechanism to check whether *pdfs* of the finite mixture distribution models for two different populations are in regular order or not by using the L_2 distance function. In the presences of non-identifiability problem, componentwise L_2 distance function should be used instead of L_2 distance function to compare *pdfs* of the finite mixture distribution models for two different populations.

REFERENCES

- McNicholas, P.D., (2016). Mixture Model-Based Classification: Chapman and Hall/CRC.
- [1] McLachlan, G. and Peel, D., (2000). Finite Mixture Models, Willey Series in Probability and Statistics. In: John Wiley & Sons, New York.
- [2] Erol, H., (2004). A Note on Non-identifiability Problem of Finite Mixture Distribution Models in Model-based Classification. Selcuk Journal of Applied Mathematics, 5(1):3-10.
- [3] Beran, R., (1977). Minimum Hellinger Distance Estimates for Parametric Models. The annals of Statistics, 5(3):445-463.
- [4] Titterington, D.M., Smith, A.F., and Makov, U.E., (1985). Statistical Analysis of Finite Mixture Distributions: Wiley.
- [5] Dempster, A.P., Laird, N.M., and Rubin, D.B., (1977). Maximum Likelihood from Incomplete Data Via the EM Algorithm. Journal of the Royal Statistical Society. Series B (methodological), pp:1-38.
- [6] Thayasivam, U., Kuruwita, C., and Ramachandran, R.P., (2015). Robust L_2E Parameter Estimation of Gaussian Mixture Models: Comparison with Expectation Maximization. 22nd International Conference, ICONIP 2015, Istanbul, Proceedings Books, pp:281-288.

-
- [7] Scott, D.W., (2001). Parametric Statistical Modeling by Minimum Integrated Square Error. *Technometrics*, 43(3):274-285. doi: 10.1198/004017001316975880.
- [8] Thayasivam, U., (2009). L_2E Estimation of Mixture Complexity. UGA.
- [9] Aitkin, M. and Rubin, D.B., (1985). Estimation and Hypothesis-Testing in Finite Mixture-Models. *Journal of the Royal Statistical Society Series B-Methodological*, 47(1):67-75.
- [10] Teicher, H., (1963). Identifiability of Finite Mixtures. *The Annals of Mathematical statistics*, 34(4):1265-1269.
- [11] Yakowitz, S.J. and Spragins, J.D., (1968). On the Identifiability of Finite Mixtures. *The annals of Mathematical statistics*, 39(1):209-214.
- [12] Kadane, J.B., (1975). The Role of Identification in Bayesian Theory. *Studies in Bayesian econometrics and statistics*.
- [13] Akdağ, S.A., (2018). Rüzgar Enerjisi Potansiyel Analizinde Karışım Dağılımları Temelli Tekniklerin Kullanılması, Doktora Tezi, İstanbul Teknik Üniversitesi, Enerji Enstitüsü.
- [14] Toher, D., Downey, G., and Murphy, T., (2005). A Comparison of Model-based and Regression Classification Techniques Applied to Near-infrared Spectroscopic data in Food Authentication Studies. Technical Report 5/10, Department of Statistics, Trinity College Dublin.
- [15] McLachlan, G.J. and Basford, K.E., (1988). *Mixture Models: Inference and Applications to Clustering*. Marcel Dekker, New York.
- [16] Fraley, C., Raftery, A.E., and Wehrens, R., (2005). Incremental Model Based Clustering for Large Data Sets with Small Clusters. *Journal of Computational and Graphical Statistics*, Vol:14, pp:1-18.
- [17] Erol, H. and Akdeniz, F., (2005). A Per-field Classification Method Based on Mixture Distribution Models and an Application to Landsat Thematic Mapper data. *International Journal of Remote Sensing* Vol:26, No:6, pp:1229-1244.
- [18] Dean, N., Murphy, T.B., and Downey, G., (2006). Updating Classification Rules with Unlabeled Data with Applications in Food Authenticity Studies. *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, Vol:55, pp:1-14.
- [19] Wehrens, R., Buydens, L., Fraley, C., and Raftery, A.E., (2004). Model Based Clustering for Image Segmentation and Large Data Sets Via Sampling. *Journal of Classification*, Vol:21, pp:231-253.
- [20] Yeung, K.Y., Fraley, C., Murua, A., Raftery, A.E. and Ruzzo, W.L., (2001). Model Based Clustering and Data Transformations for Gene Expression Data. *Bioinformatics*, 17(10):977-987.
- [21] Servi, T., (2009). Çok Değişkenli Karma Dağılım Modeline Dayalı Kümeleme Analizi, Doktora Tezi, Çukurova Üniversitesi Fen Bilimleri Enstitüsü.
- [22] Everitt, B.S. and David, J.H., (1981). *Finite Mixture Distributions. Monographs on Applied Probability and Statistics*. Chapman and Hall, London, New York.