Analysis of aggregated functional data from mixed populations with application to energy consumption

Amanda Lenzi | Camila P.E. de Souza | Ronaldo Dias | Nancy L. Garcia

Nancy E. Heckman

1Technical University of Denmark, Lyngby, Denmark
2University of British Columbia, Vancouver, British Columbia, Canada
3University of Campinas, Campinas, Brazil

Correspondence
Amanda Lenzi, Applied Mathematics and Computer Science Department, Technical University of Denmark, Richard Petersens Plads, Building 324, 2800 Kgs. Lyngby, Denmark. Email: amle@dtu.dk

Funding information
CNPq FAPESP CEMAI-CEPID CAPES CBIE-ELAP NSERC, Grant/Award Number: 442012/2014-4, 302598/2014-6, 476476/2013-5, 302956/2013-1, 2014/26419-0 and 2014/26414-9, 2013/07375-0 and 7969

Understanding energy consumption patterns of different types of consumers is essential in any planning of energy distribution. However, obtaining individual-level consumption information is often either not possible or too expensive. Therefore, we consider data from aggregations of energy use, that is, from sums of individuals’ energy use, where each individual falls into one of $C$ consumer classes. Unfortunately, the exact number of individuals of each class may be unknown due to inaccuracies in consumer registration or irregularities in consumption patterns. We develop a methodology to estimate both the expected energy use of each class as a function of time and the true number of consumers in each class. To accomplish this, we use B-splines to model both the expected consumption and the individual-level random effects. We treat the reported numbers of consumers in each category as random variables with distribution depending on the true number of consumers in each class and on the probabilities of a consumer in one class reporting as another class. We obtain maximum likelihood estimates of all parameters via a maximization algorithm. We introduce a special numerical trick for calculating the maximum likelihood estimates of the true number of consumers in each class. We apply our method to a data set and study our method via simulation.

KEYWORDS
B-splines, nonparametric regression, random effects, smoothing

1 | INTRODUCTION

Efficient distribution of energy is a problem of vital importance to electric companies around the world. An important goal in energy distribution is to not overload the energy distribution system. To prevent overload, distribution networks typically have been designed to handle the maximum demand. An alternative and more efficient strategy not only to reduce the chance of overload but also to maximize the use of existing equipment is to redistribute energy to consumers so that demand is fairly constant over time. Understanding the typical energy use pattern of each type of consumer is essential for this plan. For example, in Brazil the electrical voltage available to the vast majority of houses is either 127 V or 220 V. For this reason, households are classified as monophase (single phase/127 V) or biphasic (two phases/220 V). Usually, more modest households are monophase, suggesting that monophase and biphasic consumers have different mean electric consumption profiles. However, in both cases, residences have a spike in electricity consumption around 8 to 9 PM, due to the local habit of taking showers at night. In contrast, commercial and industrial consumers appear to have high electricity consumption between 8 AM and 6 PM.

In most regions, each energy consumer is billed according to one of several classes, such as residential (monophase or biphasic), commercial, industrial, public-use, and rural. We know that each class has a typical energy consumption pattern. For example, in Brazil, monophase residences have a lower consumption than biphasic residences. However, it is possible that a residence billed as biphasic uses energy as a monophase consumer. Also, home offices and small garage businesses are common nowadays and, although the reported class would be residential, the true consumption pattern is closer to commercial. We will call the class reported on
the electricity bill the reported class and the class that most closely indicates the consumption pattern the true class.

One way to determine consumer energy use patterns is to monitor the energy use of each consumer in a large sample composed of all types of consumers. However, obtaining consumer-level data is costly. On the other hand, aggregated data are readily available from subregions, specifically from transformers that distribute energy to a set of consumers, after decreasing voltage. Our dataset consists of a pilot study with samples of daily electric load of three transformers. See Figure 1. For instance, Figure 1(a) shows the data collected from a transformer, which distributes energy to a set of 41 consumers with fixed class memberships. The plot shows five jagged curves, the total energy consumed by the 41 consumers on each of the five weekdays in the period June 21 to June 27, 2002, inclusive. The sixth curve, which is smooth, is our estimate of the typical total energy consumption. Measurements were recorded every 15 min, in units of kilo-volt-amperes, and are plotted according to a 24-hour clock. For the data presented in Figure 1(a), we know that seven consumers reported as commercial consumers and 34 reported as residential consumers, of which five reported as monophasic and the remaining 29 as biphasic. However, as noted, these reported counts may not be accurate.

The estimation of typical load curves of electrical consumption using aggregated functional data was first done by Dias, Garcia, and Martarelli (2009) and revisited in the Bayesian framework by Dias, Garcia, and Schmidt (2013). However, these authors assumed that the reported classes were equal to the true classes. Our work is a first attempt to estimate the true class-level expected energy consumption curves from aggregated consumption data, with uncertain true class counts.

Our data set is from Companhia Paulista de Força e Luz (CPFL) Energia, a company that distributes electric energy in Southeast Brazil. The three transformers shown in Figure 1

![Figure 1](image-url)  
**FIGURE 1** Data analysis: observed data for each of the five weekdays (dashed lines) and estimated aggregated curves (solid lines) for transformers 1, 2 and 3. The horizontal axis gives the time on a 24-hour clock

| Transformer | Monophasic | Biphasic | Commercial |
|-------------|------------|----------|------------|
|             | Reported   | Estimated| Reported   | Estimated| Reported | Estimated|
| 1           | 5          | 5        | 29         | 29       | 7        | 7         |
| 2           | 4          | 4        | 43         | 43       | 3        | 3         |
| 3           | 48         | 49       | 26         | 25       | 3        | 3         |
serve 168 consumers, of whom 155 are reported as residential consumers and 13 are reported as commercial consumers.

Our goal is to use this aggregated data set along with the reported number of consumers in each class to estimate the true number of consumers in each class and to estimate each class’s expected energy consumption as a function of time. See Table 1, which contains reported counts and our estimates of true counts of consumer types for each transformer. Our analysis indicates that there is one consumer that, although reported as biphasic, has the same pattern of consumption as a monophasic residence. In our analysis, we model each consumption pattern using a linear combination of B-splines and model individual consumer perturbations from this baseline as B-splines with random coefficients. We treat the reported numbers of consumers in each class as random variables with distribution depending on the true numbers of consumers in each class and on the probabilities of a consumer in one class reporting as another class. We obtain maximum likelihood estimates of all parameters and introduce a special numerical trick for calculating the maximum likelihood estimates of the true number of consumers in each class (Theorem 1).

In the signal processing literature, the problem of aggregated information is known as linear blind signal separation (BSS). The simplest model assumes the existence of $C$ independent signals $W_1(\cdot), \ldots, W_C(\cdot)$ and the observation of at least $I$ mixtures $Y_1(\cdot), \ldots, Y_I(\cdot)$, these mixtures being $Y_i(\cdot) = \sum_{c=1}^C a_{ic} W_c(\cdot)$, with unknown coefficients $a_{i1}, \ldots, a_{ic}$. However, BSS only identifies the signals $W_1(\cdot), \ldots, W_C(\cdot)$ from the observed data, whereas in this paper, we are also interested in the mean and covariance functions of such processes. Usually, the BSS methods are based on multivariate techniques (component analysis, orthogonalization, spatio-temporal decorrelation) that consider time as a discrete set and do not take into account that the sources and the observations are continuous curves. For a review of the algorithms and the statistical principles of BSS, see Cardoso (1998), Choi, Cichocki, Park, and Lee (2005), and Comon and Jutten (2010).

This paper is organized as follows. Details of our model are in Section 2, with estimates described in Section 3. In Section 4, we give details of implementation as well as extend our model and estimation procedure to the case where there are replicate observations from each transformer. Section 5 contains the analysis of the energy consumption data for the three transformers. Section 6 contains the results of a simulation study.

2 | NOTATION AND MODEL

We index transformer by $i = 1, \ldots, I$, class by $c = 1, \ldots, C$, and consumer served by transformer $i$ by $l = 1, \ldots, N_i$. We denote the true class of individual $l$ served by transformer $i$ as $c_{li}$ and the reported class as $r_{li}$. We assume that the $c_{li}$’s and $r_{li}$’s take values in the same set and do not change over the time period of the data collection. We let $W_{li}(t)$ equal the electricity consumption at time $t$ of individual $l$ served by transformer $i$. We model $W_{li}$ as a hidden, random process whose distribution depends on the value of $c_{li}$. In transformer $i$, we do not observe the $W_{li}$’s but rather their sum plus measurement error, at $n_i$ time points, $t_1 < t_2 < \cdots < t_{n_i}$. For simplicity of notation and exposition, we only consider the case with $n_i \equiv n$ and $t_{ij} \equiv t_j$ for all $i$ and for $j = 1, \ldots, n$. We do not observe $M_{ci}$, the true number of consumers in class $c$ served by transformer $i$. Rather, we observe $R_{ci}$, the number of reported consumers in class $c$ served by transformer $i$. Throughout, we assume that random quantities are independent from transformer to transformer.

We are interested in estimating the true counts of consumers in each class, served by each transformer, and the typical usage curve of a consumer of class $c$, which is simply the expected value of $W_{li}(t)$ when $c_{li} = c$.

2.1 | Model for $W_{li}$ and the observed aggregated electricity consumption

We suppose that $W_{li}$, the energy consumption of consumer $l$ served by transformer $i$ of consumer type $c_{li} = c$, is given by

$$W_{li}(t) \bigg|_{c_{li}=c} = a_c(t) + a^*_{ci}(t)$$

where $a_c$ is the non-random typical usage curve (also called the typology) in class $c$ and $a^*_{ci}$ is the consumer-specific random perturbation. We model $a_c$ and $a^*_{ci}$ with B-splines basis functions $\phi_1, \ldots, \phi_K$ and $\psi_1, \ldots, \psi_K$, respectively. Letting $\phi(t) = (\phi_1(t), \ldots, \phi_K(t))$ and $\psi(t)$ defined similarly,

$$a_c(t) = \sum_{k=1}^K \gamma^c[k] \phi_k(t) \equiv \psi(t)\gamma^c$$

and

$$a^*_{ci}(t) = \sum_{k=1}^{K^*} \gamma^{ci}[k] \psi_k(t) \equiv \psi(t)\gamma^{ci}.$$

We suppose that the vector $\gamma^c$ is an unknown parameter vector and the vectors $\gamma^{ci}$ are random effects, normally distributed, independent with mean 0 and covariance matrix $\Sigma_c$. Note that this implies that, given $c_{li} = c$, $W_{li}$ is a Gaussian process with mean $a_c$ and covariance function

$$\sigma(s, t) \equiv \text{cov}(W_{li}(s), W_{li}(t)) = \text{cov}(a^*_c(s), a^*_c(t)) = \psi(s)\Sigma_c \psi(t).$$

The process $a^*_{ci}$ allows us to account for within consumer correlation over time.

Notice that the vectors of basis function evaluations, $\phi(t)$ and $\psi(t)$ in (1) and (2), respectively, do not depend on the consumer type $c$ – that is, we use the same number of basis functions ($K$ and $K^*$) and the same knot locations for all types of consumers. In our application and simulation studies, we
choose the number of basis functions by eye. One could use a different number of basis functions for each class, choosing the number of basis functions automatically using, for instance, the algorithms proposed by Luo and Wahba (1997), Dias (1998), Bodin, Villemoes, and Wahlberg (2000), or De Vore, Petrova, and Temlyakov (2003).

We observe the data vector \( Y_i = (Y_i(t_1), \ldots, Y_i(t_n)) \) where, for \( \varepsilon_i(\cdot) \) and \( W_i(\cdot) \) independent, with \( \varepsilon_i \) Gaussian white noise with \( \text{var}(\varepsilon_i(\cdot)) = \sigma^2 \),

\[
Y_i(t) = \sum_{c=1}^{C} \sum_{l=1}^{N_i} W_{il}(t) I\{c_{il} = c\} + \varepsilon_i(t)
\]

(3)

\[
= \sum_{c=1}^{C} M_i \alpha_i(t) + \sum_{l=1}^{N_i} \sum_{c=1}^{C} \alpha_{cilt}(t) I\{c_{il} = c\} + \varepsilon_i(t).
\]

We easily see that \( \text{E}(Y_i(t)) = \sum_{c=1}^{C} M_i \alpha_i(t) \) and, because of the independence of the \( \alpha_{cit} \)'s,

\[
\text{cov}(Y_i(s), Y_i(t)) = \sum_{c=1}^{C} \sum_{l=1}^{N_i} \text{I}(c_{il} = c) \text{cov}(\alpha_{cilt}(s), \alpha_{cilt}(t))
\]

\[+ \text{cov}(\varepsilon_i(s), \varepsilon_i(t)).
\]

Defining the \( n \) by \( K \) matrices \( \Phi \) and \( \Psi \) as

\[
\Phi[j, k] = \varphi_k(t_j) \quad \text{and} \quad \Psi[j, k] = \psi_k(t_j)
\]

yields

\[
Y_i \sim \mathcal{N}\left( \Phi \sum_{c=1}^{C} M_c \gamma^c, \Psi \sum_{c=1}^{C} M_c \Sigma_c^\gamma \Psi' + \sigma^2 I \right).
\]

(4)

From (4), we see that, without further information about the \( M_c \)'s, the distribution of \( Y_i \) is not identifiable – there are an infinite number of distinct \( M_c \)'s, \( \gamma^c \)'s, and \( \Sigma_c^\gamma \)'s yielding the same distribution. However, when we observe the reported counts, the joint distribution of \( Y_i \) and \( R_{1i}, \ldots, R_{Ci} \) is identifiable, provided we know the rates of misreporting.

### 2.2 Model for the reported counts of consumer classes

Recall that \( r_{lj} \) is the class reported for consumer \( l \) served by transformer \( i \), that \( c_{il} \) is the consumer’s true class, that \( M_c \) is the true number of consumers of class \( c \) served by transformer \( i \) and that \( R_{ci} \) is the corresponding reported number. We suppose that all consumers report some class, that is, that \( N_i \), the total number of consumers served by transformer \( i \), is equal to \( \sum_{c=1}^{C} R_{ci} \), which is also equal to \( \sum_{c=1}^{C} M_c \). Recall that, in our model, the \( M_c \)'s are fixed parameters and the \( R_{ci} \)'s are random variables. We require a model for consumer reporting, that is, a model for \( R_{1i}, \ldots, R_{Ci} \) when the true counts are \( M_{1i}, \ldots, M_{Ci} \).

To model consumer reporting, we propose that there is a known matrix of reporting probabilities \( \mathcal{P} \), not depending on the transformer, with \( \mathcal{P}(c, r) = P(r_{lj} = r|c_{il} = c) \), \( r, c = 1, \ldots, C \), the probability that a consumer of class \( c \) reports as being of class \( r \). We assume that consumers report independently of other consumers.

For convenience, we drop the transformer subscript \( i \) and let \( x_{cj} \) be the number of consumers who are of class \( c \) but report they are of class \( j \). Then the reported number of consumers of class \( j \) is \( R_j = \sum x_{cj} \). The true number of consumers of class \( c \) is \( M_c = \sum x_{cj} \).

For \( c = 1, \ldots, C \), let \( X_c \equiv (x_{c1}, \ldots, x_{cC})' \), the vector of reported counts from consumers of class \( c \). Then \( X_c \) is multinomial \( (M_c, \mathcal{P}(c, 1), \ldots, \mathcal{P}(c, C)) \). By independence of consumer reporting, \( X_{11}, \ldots, X_c \) are independent. Thus, we have defined the joint distribution of \( R_l = \sum x_{c1}, \ldots, R_C = \sum x_{cC} \) when the true class counts are \( M_1, \ldots, M_C \).

### 2.3 Extending to replicates

Equation (3) considers the case where there is only one \( W_{il} \) for each consumer and can be used to model the situation where \( W_{il} \) is the power usage on a specified date or the average of usages on a sequence of dates. We can easily extend (3) to model the situation where there are \( D_l \) replicates of the \( W_{il} \)'s for consumer \( l \) served by transformer \( i \).

In this case, our observations from transformer \( i \) consist of \( Y_{id} \equiv (Y_{id}(t_1), \ldots, Y_{id}(t_n))' \), \( d = 1, \ldots, D_l \), and the aggregated model can be written as (cf. (3))

\[
Y_{id}(t) = \sum_{l=1}^{N_i} W_{lid}(t) + \varepsilon_{id}(t)
\]

\[= \sum_{l=1}^{N_i} \sum_{c=1}^{C} W_{lid}(t) I\{c_{il} = c\} + \varepsilon_{id}(t)
\]

\[= \sum_{c=1}^{C} M_c \alpha_c(t) + \sum_{l=1}^{N_i} \sum_{c=1}^{C} \alpha_{c lid}(t) I\{c_{il} = c\} + \varepsilon_{id}(t)
\]

(5)

for \( \varepsilon_{id}(\cdot) \) and \( W_{lid}(\cdot) \) independent, with \( \varepsilon_{id} \) independent Gaussian white noise with \( \text{var}(\varepsilon_{id}(t)) = \sigma^2 \).

We use this extension in the simulation study and in the data analysis to model a consumer’s energy consumption on a sequence of 5 days, considering these 5 days as five replicates. The calculation and maximization of the likelihood for the replicate case is a straightforward modification of the iterative procedure described in Section 3.

### 3 Maximum likelihood estimation

We partition the parameters as follows. Let \( G = \{\gamma^1, \ldots, \gamma^C\} \) be the set of parameter vectors defining the class means, the \( \alpha_c \)'s, and let \( S = \{\Sigma_1, \ldots, \Sigma_C\} \) be the set of parameters defining the class variance structure. Recall that \( \sigma^2 \) is the measurement error variance. Let \( M_i = (M_{1i}, \ldots, M_{Ci})' \) be the vector of true class counts in transformer \( i \) and \( M \) the collection of true counts, \( M_1, \ldots, M_l \).

The data are \( \mathbf{Y}_i \) as in (4) and \( \mathbf{R}_i = (R_{i1}, \ldots, R_{Ci})' \), the reported counts, for transformers \( i = 1, \ldots, I \).

By the independence of the transformers, the log likelihood is

\[
\mathcal{L}(\mathcal{G}, S, \sigma^2, \mathbf{M}) = \sum_{i=1}^{I} \mathcal{L}_i(\mathcal{G}, S, \sigma^2, \mathbf{M}_i),
\]

where

\[
\mathcal{L}_i(\mathcal{G}, S, \sigma^2, \mathbf{M}_i) = \mathcal{L}_i(\mathcal{G}, S, \sigma^2, \mathbf{M}_i \mid \mathbf{Y}_i, \mathbf{R}_i)
\]

\[
= \log \left[ \prod_{c=1}^{C} \mathcal{L}_i(\mathcal{G}, S, \sigma^2, c, \mathbf{M}_i) \times \mathcal{P}(\mathbf{R}_i \mid \mathbf{M}_i) \right].
\]

The equality in (7) comes from the fact that the distribution of \( \mathbf{Y}_i \) depends only on \( \mathcal{G}, S, \sigma^2 \) and on \( \mathbf{M}_i \), the true class counts, and not on \( \mathbf{R}_i \), the reported counts. The expression for \( \log f_{\mathcal{Y}_i}(\mathbf{Y}_i \mid \mathcal{G}, S, \sigma^2, \mathbf{M}_i) \) follows directly from our model (4). The second term, \( \mathcal{P}(\mathbf{R}_i \mid \mathbf{M}_i) \), is the probability mass function of \( \mathbf{R}_i \) and is discussed in detail in Sections 3.3 and 3.4.

\[\hat{I}_1 \equiv \sum_{i=1}^{I} \log \left[ \Psi \sum_{c=1}^{C} M_{ci} \Sigma_c \Psi' + \sigma^2 \right] + \sum_{i=1}^{I} \left( Y_i - \sum_{c=1}^{C} M_{ci} \phi_c \right)' \left( \Psi \sum_{c=1}^{C} M_{ci} \Sigma_c \Psi' + \sigma^2 \right)^{-1} \left( Y_i - \sum_{c=1}^{C} M_{ci} \phi_c \right) \]

3.1 Maximization procedure

We carry out this maximization iteratively and step-wise using a block coordinate ascent method such as the one described and studied by Tseng (2001). Finding the estimates for the true class counts is not straightforward; details are given in Section 3.3. Our proposed maximization algorithm is described as follows.

At the \( s \)th iteration, we update the parameter estimates \( \mathcal{G}^{(s)}, S^{(s)}, \sigma^{(s)}, \phi^{(s)} \) and \( M^{(s)} \) to \( \mathcal{G}^{(s+1)}, S^{(s+1)}, \sigma^{(s+1)}, \phi^{(s+1)} \) and \( M^{(s+1)} \) so that

\[
\mathcal{L} \left( \mathcal{G}^{(s+1)}, S^{(s+1)}, \sigma^{(s+1)}, M^{(s+1)} \right) \geq \mathcal{L} \left( \mathcal{G}^{(s)}, S^{(s)}, \sigma^{(s)}, M^{(s)} \right).
\]

We initialize the procedure by taking \( M^{(0)} = \mathbf{R}_i \). We then carry out the following two steps until convergence. Details of each step follow in Sections 3.2, 3.3, and 4.

1. Given \( M^{(s)} \), we let \( \mathcal{G}^{(s+1)}, S^{(s+1)}, \sigma^{(s+1)} \) maximize the log likelihood, or at least not decrease the log likelihood.
2. Given \( \mathcal{G}^{(s+1)}, S^{(s+1)}, \sigma^{(s+1)} \), we let \( M^{(s+1)} \) maximize the log likelihood, or at least not decrease the log likelihood.

Our iterative procedure has not shown any convergence problems so far: in all of our analyses — of simulated data and of the transformer data — the procedure always converged. To study the possible problem of multimodality of the log likelihood function, for a few data sets, we used several different sets of starting values for the parameter estimates. In all cases, the algorithm converged to the same final parameter estimates.

3.2 Step 1: updating \( \mathcal{G}, S \) and \( \sigma^2 \)

To update our estimates of \( \mathcal{G}, S \), and \( \sigma^2 \), we maximize the log likelihood as a function of \( \mathcal{G}, S, \sigma^2 \) keeping \( M \), the true class counts, fixed. From (6) and (7), we can rewrite the log likelihood as

\[
\mathcal{L}(\mathcal{G}, S, \sigma^2, \mathbf{M}) = \sum_{i=1}^{I} \log f_{\mathcal{Y}_i}(\mathbf{Y}_i \mid \mathcal{G}, S, \sigma^2, \mathbf{M}_i) + \sum_{i=1}^{I} \log \mathcal{P}(\mathbf{R}_i \mid \mathbf{M}_i).
\]

Because the expression for \( \mathcal{P}(\mathbf{R}_i \mid \mathbf{M}_i) \) does not depend on \( \mathcal{G}, S, \sigma^2 \), the \( \mathcal{G}, S, \sigma^2 \) that maximize (8) are the \( \mathcal{G}, S, \sigma^2 \) that maximize \( \hat{I}_1 \) or minimize \( -2\hat{I}_1 \). Thus, using \( \mathbf{Y}_i \)'s normal distribution in (4) and disregarding the constant terms, we see that we must minimize

\[\mathcal{L}(\mathcal{G}, S, \sigma^2, \mathbf{M}) = \sum_{i=1}^{I} \log f_{\mathcal{Y}_i}(\mathbf{Y}_i \mid \mathcal{G}, S, \sigma^2, \mathbf{M}_i) + \sum_{i=1}^{I} \log \mathcal{P}(\mathbf{R}_i \mid \mathbf{M}_i).
\]

3.3 Step 2: updating \( M_1, \ldots, M_I \)

Note that, to maximize the log likelihood with respect to \( M_1, \ldots, M_I \), we can carry out \( I \) separate maximizations, one for each transformer. That is, for each fixed \( i = 1, \ldots, I \), we seek \( M_{ci}, c = 1, \ldots, C \) to maximize \( \mathcal{L}(\mathcal{G}, S, \sigma^2, \mathbf{M}_i) \) in (7), treating \( \mathcal{G}, S, \sigma^2 \) as fixed.

Maximizing \( \mathcal{L}_i \) as a function of \( M \) brings challenges. The function \( \mathcal{P}(\mathbf{R}_i \mid \mathbf{M}_i) \) has a complex form. Furthermore, Newton–Raphson type methods of maximization are inappropriate because the possible values of \( M_{ci} \) are integers.
with plausible values typically in a small range near the reported counts.

We can approximate the function \( P(\mathbf{R}_i|\mathbf{M}_i) \) via simulation. The most natural simulation is a “brute force” one: for each fixed possible value of \( \mathbf{M}_i \) we would generate a large number of \( \mathbf{R}_i \)’s according to the model described in Section 2.2 and calculate the proportion of times the generated \( \mathbf{R}_i \) is equal to the observed \( \mathbf{R}_i \). As we are maximizing \( \mathcal{L}_i \) with respect to \( \mathbf{M}_i \), we would have to repeat the simulation and calculation for all plausible values of \( \mathbf{M}_i \). Even if we can restrict simulations to a small range of plausible values of \( \mathbf{M}_i \), this procedure still requires many simulations per transformer \( i \).

\[
P(\mathbf{R} = \mathbf{r}|\mathbf{M} = \mathbf{m}) = P(R_1 = r_1, \ldots, R_C = r_C|\mathbf{M}_1 = m_1, \ldots, \mathbf{M}_C = m_C)
\]

Thus, the entries of each \( \tilde{X}_j \) sum to \( r_j \) and the multinomial probability associated with the \( c \)th component of \( \tilde{X}_j \) is proportional to \( P(c,j) \), the probability that a consumer of class \( c \) reports being in class \( j \). So \( \tilde{X}_j \) can be considered as the vector of counts, dividing up the \( r_j \) consumers who have reported being class \( j \) into their true classes. The distribution of \( \tilde{X}_1, \ldots, \tilde{X}_C \) relates to the distribution of \( R_1, \ldots, R_C \), as given in the following theorem.

**Theorem 1.** For the random variables \( R_1, \ldots, R_C \) as defined in Section 2.2 and for \( \tilde{X}_1, \ldots, \tilde{X}_C \) independent with distributions as defined above,

\[
H(m_1, \ldots, m_C) = P \left\{ \sum_{j=1}^C \tilde{X}_j[c] = m_c, c = 1, \ldots, C \right\}.
\]

The proof of Theorem 1 is presented in the Appendix.

To estimate \( H \) in (11), we simulate \( B \) independent sets distributed as \( \{\tilde{X}_1, \ldots, \tilde{X}_C\} \), with the \( b \)th simulated data set denoted \( \{\tilde{X}_{1b}, \ldots, \tilde{X}_{Cb}\} \), \( b = 1, \ldots, B \). We let

\[
\hat{H}(m_1, \ldots, m_C) = \frac{1}{B} \sum_{b=1}^B \left\{ \sum_{j=1}^C \tilde{X}_{jb}[c] = m_c, c = 1, \ldots, C \right\}.
\]

See Section 1 of the Supplementary Material for a simple example.

### 3.4 Alternate form for \( P(\mathbf{R}_i|\mathbf{M}_i) \)

For convenience, we will once again drop \( i \), the subscript indicating the transformer. Here, we derive an expression for the probability that \( R_1 = r_1, \ldots, R_C = r_C \) in terms of random vectors, \( \tilde{X}_1, \ldots, \tilde{X}_C \): for \( j = 1, \ldots, C \), let \( \tilde{X}_j \) be multinomial with parameters \( r_j, p_{1j}, p_{2j}, \ldots, p_{cj} \) with

\[
p_{cj} = \frac{P(c,j)}{\sum_{l=1}^C P(l,j)}.
\]

### 4 DETAILS OF IMPLEMENTATION

We now give details of implementation concerning choice of starting values of the parameter estimates, maximization of the log likelihood under the restriction that \( \Sigma_c = \sigma_{r,c}^2 I \) and the choice of software.

As previously mentioned, we recommend using the reported counts as the initial estimates of the true counts.
To obtain the initial estimate of $\sigma^2$, we need to pool across transformers.

Because we assume that the $\gamma^{dci}$s have covariance $\Sigma_c = \sigma^2_{r,c} I$, the set of covariance parameters $\Sigma$ is equal to $\{\sigma^2_{r,c}, c = 1, \ldots, C\}$. We use method of moments for our initial estimates of the $\sigma^2_{r,c}$s, using the fact that $Y_i$ has a multivariate distribution with covariance matrix $\left( \sum_{c=1}^C M_c \sigma^2_{r,c} \right) \Psi \Psi' + \sigma^2 I$. For calculating the initial estimates, we suppose that prior knowledge tells us that $\sigma^2_{r,c} = s_c \sigma^2_{r,c}$ for some known $s_c$’s, $c = 1, \ldots, C - 1$. To extend this notation, we set $s_C = 1$.

Write

$$\sum_{i,j} \text{var}(Y_{i,j}) = \sigma^2_{r,C} \sum_i \left[ \sum_{c=1}^C M_c s_c \right] \text{trace}(\Psi \Psi') + \text{In}\sigma^2.$$  \hspace{1cm} (12)

We use our initial estimates of $M_c$ and $\gamma^c$ to form the estimate of the left side of (12): $\sum_{i,j} \text{var}(Y_{i,j}) = ||Y_i - \hat{M}_{ci} \Phi \gamma^c||^2$. We substitute our initial estimates of the $M_c$’s and $\sigma^2$’s into the right side of (12) and then solve for our estimate of $\sigma^2_{r,C}$. This yields our initial estimate, $\hat{\sigma}^{(0)}_{r,C}$, and our other initial estimates, $\hat{\sigma}^{(0)}_{r,c}$.

To maximize the likelihood, we iterate through Steps a, b and c of Section 3.2 and the procedure of updating the $M$’s in Section 3.3. Step a is straightforward. For the one-dimensional minimization of Step b, that is, for updating estimates of $\sigma^2$, we use the $R$ function optimize.

For Step c of the updating algorithm, we hold $G$ and $\sigma^2$ fixed and minimize $l_i$ with respect to the $\sigma^2_{r,c}$’s. First, write $\Psi \sum_{c=1}^C M_c \sigma^2_{r,c} \Psi' + \sigma^2 I = \left( \sum_{c=1}^C M_c \sigma^2_{r,c} \right) \Psi \Psi' + \sigma^2 I$. Writing the eigenvalue–eigenvector decomposition of $\Psi \Psi'$ as $Q \Gamma Q'$ with $Q$ diagonal and $Q$ orthonormal yields

$$\left[ \sum_{c=1}^C M_c \sigma^2_{r,c} \Psi' \Psi + \sigma^2 I \right]^{-1} = Q' \left[ \sum_{c=1}^C M_c \sigma^2_{r,c} \Gamma + \sigma^2 \right]^{-1} Q.$$  \hspace{1cm} \text{(12)}

Let $\Delta_i = \Delta_i \left( \sigma^2_{r,1}, \ldots, \sigma^2_{r,C} \right)$ be the diagonal matrix $\Delta_i = \left( \sum_{c=1}^C M_c \sigma^2_{r,c} \right) \Gamma + \sigma^2 I$. Thus, letting $Y_i^s = Q(Y_i - \sum_{c} M_c \Phi \gamma^c)$, we must find $\sigma^2_{r,1}, \ldots, \sigma^2_{r,C}$ to minimize

$$l_i \left( \gamma^c, \left\{ \sigma^2_{r,1}, \ldots, \sigma^2_{r,C} \right\}, \sigma^2 \right) = \sum_l \log|\Delta_l| + \sum_l Y_i^s \gamma^c \Delta_l^{-1} Y_i^s.$$  \hspace{1cm} \text{(12)}

Because $\Delta_i$ is diagonal, the minimization can be easily carried out numerically via a $C$-dimensional optimization. Here, we use the $R$ function optim with method L-BFGS-B, which is a modification of the quasi-Newton method. This function allows specification of a lower and upper bound for each variable, which we need to force variance parameter estimates to be non-negative. Note that we can calculate $Q$ and $\Gamma$ at the beginning of our iterations, as they are determined by the choice of basis.

**Extending to replicates.** To obtain the initial estimate of $\sigma^2$ in the replicate case, we first fit a smoothing spline curve to each replicate in transformer $i$ and calculate the sample variance of the residuals of that fit adjusting for the correct degrees of freedom, which are based on the trace of the smoothing hat matrix. We then pool those variances across replicates and transformers to obtain our initial estimate of $\sigma^2$. More details on smoothing-based estimation of variances and calculation of appropriate degrees of freedom can be found in Wahba (1983).

In the case where we observe $D$ replicates from transformer $i$, to compute $\hat{\sigma}^{(0)}_{r,c}$, and our other initial estimates, $\hat{\sigma}^{(0)}_{r,c}$, we modify the calculations, summing both sides of (12):

$$\sum_{d,i,j} \text{var}(Y_{i,d,j}) = D \left\{ \sum_{c=1}^C M_c s_c \right\} \text{trace}(\Psi \Psi') + \text{In}\sigma^2,$$

and estimating the variance of $Y_{i,d,j}$ by the sample variance of $Y_{i,d,j}$.

The modification of Step c for the replicate case is straightforward.

**5 | DATA ANALYSIS**

Recall that our data set consists of a pilot study with energy consumption data from three transformers, with consumption recorded every 15 min during 5 days of a particular week. For each transformer, we have $C = 3$ reported consumer types, residential monophasic ($c = 1$), residential biphasic ($c = 2$) and commercial ($c = 3$). Also, we assume that the true consumption belongs to one of these three reported classes and that the sum of reported counts equals the sum of true counts.

In our analysis, we use the methods of Section 4 with $D = 5$ replicates, one for each weekday. In addition, in order to ensure that $\hat{\alpha}_c(t)$ is always positive, we take the approach presented by Ramsay & Silverman (2005) and write

$$\hat{\alpha}_c(t) = \exp \left( \sum_{k=1}^K \gamma^{sc}[k] \varphi_k(t) \right)$$

$$= \exp \left( \varphi(t)^\gamma^{sc} \right) \text{ for } c = 1, \ldots, C.$$

We then use a numerical optimization method to obtain $\gamma^{sc}[k] = \left( \gamma^{s1[k+1]}, \ldots, \gamma^{sC[k+1]} \right)'$ in Step 1 of our maximization procedure described in Section 3.2.

In the analysis, we consider the matrix of reporting probabilities given by

$$\mathcal{P} = \begin{bmatrix} 0.96 & 0.02 & 0.02 \\ 0.98 & 0.02 & 0.02 \\ 0.05 & 0.05 & 0.9 \end{bmatrix},$$

where, for instance, a commercial consumer is reported as either a monophasic or biphasic residential consumer, each with probability 0.05. This misreporting may occur because a
small business opened in a former monophasic/biphasic residence. We also note in (13) that a biphasic consumer cannot be reported as a monophasic because the energy company knows the voltage power of the residences. Of course, the entries in this matrix need to be obtained from the experts in the field. In the simulation studies and in the analysis of our energy consumption dataset, we verified that the results were not impacted by slight changes in the assumed $P$ matrix.

It is well known, in regression splines, that determining the number of knots and their placement is crucial. A badly placed knot vector can either cause unwanted multimodality or fail to “catch a bump”. Based on previous works, Dias et al. (2009) and Dias et al. (2013), in the estimation procedure we consider the same $\varphi$’s as $\psi$’s, namely a set of twelve cubic B-spline basis functions with equally spaced knots.

We consider the case where $\Sigma_c = \sigma^2_{c,1} I_c$, $c = 1, 2, 3$. We find initial estimates of $\sigma^2$, $\sigma^2_{r,1}$, $\sigma^2_{r,2}$ and $\sigma^2_{r,3}$ using the methods described in Section 4. After our iterative maximization of the likelihood, we obtain our final estimates for the variance parameters: $\hat{\sigma}^2 = 7.74$, $\hat{\sigma}^2_{r,1} = 0.089$, $\hat{\sigma}^2_{r,2} = 0.019$ and $\hat{\sigma}^2_{r,3} = 0.730$.

Table 1 presents the number of reported consumers from the residential and commercial classes and our estimates of the true numbers of consumers. According to our estimates, Transformers 1 and 2 have the correct number of reported consumers for all classes, while Transformer 3 has one residence receiving two voltages (127/220V) but with pattern usage of a monophasic residence.

The estimated typical energy usage curves of residential and commercial consumers are shown in Figure 2. Panel (d) shows the three estimates together, while panels (a), (b), and (c) show each estimate separately. We can see that, most of the time, the residential biphasic usage is higher than the monophasic residential usage. Commercial usage drops close to zero around 5 AM. The peak load of commercial consumers is almost 5 times the peak load of monophasic residential consumers and 3 times that of biphasic residential consumers.

Residential consumers have a high peak of energy consumption around 8–9 PM, due to the local habit of taking showers at night. There is another peak around noon: if it is “real”, it probably occurs because Brazilians return home for lunch. Commercial consumers have a load that only increases between 12 PM (lunch time) until a peak at around 6 PM.

Figure 1 shows the estimated aggregated curves $\left(\sum_{c=1}^{3} \hat{M}_c \hat{a}_c(t)\right)$ for each transformer along with the observed electrical load.

6 | SIMULATION STUDIES

For simplicity, in the simulation studies, we continue using the same nomenclature and notation as in the real dataset,
referring to aggregated curves as transformer loads, considering three classes referred to as consumer classes: monophasic (class \(c = 1\)), biphasic (class \(c = 2\)), and commercial (class \(c = 3\)), and so on.

We study two scenarios for the typical consumer usage curves (\(\alpha_1\), \(\alpha_2\), and \(\alpha_3\)) and the reported counts (\(M_1\), \(M_2\), and \(M_3\)).

**Case 1.** The three functions, \(\alpha_1\), \(\alpha_2\), and \(\alpha_3\), are of the same scale and the \(M_1\)’s, \(M_2\)’s, and \(M_3\)’s are balanced, that is, are approximately equal.

**Case 2.** The function \(\alpha_1\) is of a smaller scale than \(\alpha_2\) which is of a smaller scale than \(\alpha_3\). The \(M_1\)’s and \(M_2\)’s are larger than the \(M_3\)’s.

For each case, we carry out two simulation studies, one with data from three transformers and one with data from thirty transformers. For each of the resulting four simulation studies, we generate 200 data sets. All four simulation studies contain 5 days (replicates) of power consumption curves for each transformer. For simplicity, we generate the days independently, although we note that consumer level day-to-day usages are probably correlated.

Details of the scenarios and data generation are given in Section 6.1. Results of the simulation studies are given in Section 6.2. Further simulation studies are presented in the Supplementary Material.

### 6.1 Data generation

In each of our four simulation studies, energy consumption for each transformer is “observed” every 15 min, so that there are 96 measurements taken each day, with time \(t \in (0,24]\) hours. All data sets are generated according to the model in (5) with \(\widehat{\mathbf{Y}}_{id}\), the data vector of observations from transformer \(i\) on day \(d\), distributed as in (4).

![Figure 3](image-url)

**FIGURE 3** Simulation study Case 1 (\(\alpha_1\), \(\alpha_2\) and \(\alpha_3\) are of the same scale and the \(M\)’s are balanced): first and third quartiles of the 200 estimated typologies for Class 1 (monophasic), Class 2 (biphasic) and Class 3 (commercial). (a) Data generated with 3 transformers and 5 replicates. (b) Data generated with 30 transformers and 5 replicates. The solid curve is the true typology used to generate the data, the shaded gray area corresponds to the area between the first and third quartiles.
FIGURE 4  Simulation study Case 2 (\( \alpha_1 \) is of a smaller scale than \( \alpha_2 \), which is of a smaller scale than \( \alpha_3 \). The \( M_1 \)'s and \( M_2 \)'s are larger than the \( M_3 \)'s): first and third quartiles of the 200 estimated typologies for Class 1 (monophasic), Class 2 (biphasic) and Class 3 (commercial). (a) Data generated with 3 transformers and 5 replicates. (b) Data generated with 30 transformers and 5 replicates. The solid curve is the true typology used to generate the data, the shaded gray area corresponds to the area between the first and third quartiles.

FIGURE 5  Simulation study Case 2 (\( \alpha_1 \) is of a smaller scale than \( \alpha_2 \), which is of a smaller scale than \( \alpha_3 \), the \( M_1 \)'s and \( M_2 \)'s are larger than the \( M_3 \)'s) with three transformers and five replicates: estimate of the total electric load (\( \hat{M}_1 \hat{\alpha}_1(t) + \hat{M}_2 \hat{\alpha}_2(t) + \hat{M}_3 \hat{\alpha}_3(t) \)) for one of the three transformers.

We set \( \sigma^2 = 3.5 \) and we assume that the \( \Sigma_e \)'s are diagonal, with \( \Sigma_e = \sigma^2_{\varepsilon,\varepsilon} I \). For the basis functions, we use the same \( \varphi \)'s as \( \psi \)'s, a set of 12 cubic B-splines with equally spaced knots.

We obtain \( \alpha_1, \alpha_2, \) and \( \alpha_3 \) along with the corresponding \( \gamma_1, \gamma_2, \) and \( \gamma_3 \) for the two cases as follows. For Case 2, we fit the proposed model to the data. For Case 1, we rescale the Case 2 \( \alpha_1, \alpha_2, \) and \( \alpha_3 \) by rescaling the associated \( \gamma_1, \gamma_2, \) and \( \gamma_3 \) so that all components are between 0 and 1. For instance, to rescale \( \alpha_1 \), let \( a_1 \) equal the minimum of \( \gamma_1 \)'s components and \( b_1 \) equal the maximum. We define the rescaled \( \alpha_1(t) \) as \( \varphi(t) \gamma_1 \) with \( \gamma_1[k] = (\hat{\gamma}_1[k] - a_1)/(b_1 - a_1) \). The resulting functions are plotted in Figures 3 (Case 1) and 4 (Case 2).

For Case 2, we set the variance parameters for the consumer-level energy consumption curves equal to the estimates given by the fit of the proposed model to the data: \( \sigma^2_{\gamma,1} = 0.089, \sigma^2_{\gamma,2} = 0.019 \) and \( \sigma^2_{\gamma,3} = 0.73 \). For Case 1, to maintain the relative variability of the consumer-level curves about the \( \alpha \)'s, we rescale \( \sigma^2_{\varepsilon,\varepsilon} \) by dividing it by the appropriate constant:
We determine the true consumer class counts within each transformer as follows. In Case 1, each transformer serves 75 consumers, and we force the three classes to contain the same total number of consumers, totalled across all transformers. We then divide this total number at random among the transformers. For instance, for the simulations with data generated from three transformers with 225 consumers, 75 consumers are commercial, 75 are monophasic residential and 75 are biphasic residential. We randomly assign each of the 225 consumers to one of the three transformers. In Case 2, when we have three transformers, we use the estimated counts from our data analysis as the true class counts. When we have 30 transformers, we re-use these true class counts for the first three transformers. In the remaining 27 transformers, for Case 2, to maintain the variability among the class counts that is observed in the data set, we set the true number of consumers in the three classes to rounded realizations of independent normal random variables with means equal to the counts in the first three transformers. For instance, to generate the monophasic class counts in the remaining 27 transformers, we set \( \mu_{1i} \) equal to the estimated monophasic counts in transformer \( i, i = 1, 2, 3 \). From Table 1, we see that \( \mu_{11} = 5, \mu_{21} = 4, \) and \( \mu_{31} = 49 \). We generate nine independent normal random variables with mean \( \mu_{11} \) and variance \( \mu_{11}/3 \). We round these to the nearest integer to produce the true monophasic class counts of the first 9 of the 27 transformers. We repeat this procedure for the counts of biphasic and commercial classes, defining the \( \mu \)'s accordingly. Then, for each simulated data set, we generate the reported counts for the \( i \)th transformer by generating multinomial vectors \( X_{i1}, X_{i2}, \) and \( X_{i3} \) as described in Section 2.2, using the matrix of reporting probabilities as in (13).

\[
(b_i - a_i)^2, \ c = 1, 2, 3, \ \text{yielding} \ \sigma_{r,1}^2 = 0.090, \ \sigma_{r,2}^2 = 0.007, \ \text{and} \ \sigma_{r,c}^2 = 0.025.
\]

We have 30 transformers, we re-use these true class counts for the first three transformers. In the remaining 27 transformers, for Case 2, to maintain the variability among the class counts that is observed in the data set, we set the true number of consumers in the three classes to rounded realizations of independent normal random variables with means equal to the counts in the first three transformers. For instance, to generate the monophasic class counts in the remaining 27 transformers, we set \( \mu_{1i} \) equal to the estimated monophasic counts in transformer \( i, i = 1, 2, 3 \). From Table 1, we see that \( \mu_{11} = 5, \mu_{21} = 4, \) and \( \mu_{31} = 49 \). We generate nine independent normal random variables with mean \( \mu_{11} \) and variance \( \mu_{11}/3 \). We round these to the nearest integer to produce the true monophasic class counts of the first 9 of the 27 transformers. We repeat this procedure for the next 9 of the 27 transformers, using \( \mu_{21} \). We repeat once more for the remaining 9 of the 27 transformers, using \( \mu_{31} \). We carry out this procedure for the counts of biphasic and commercial classes, defining the \( \mu \)'s accordingly. Then, for each simulated data set, we generate the reported counts for the \( i \)th transformer by generating multinomial vectors \( X_{i1}, X_{i2}, \) and \( X_{i3} \) as described in Section 2.2, using the matrix of reporting probabilities as in (13).
6.2 | Analysis and results

In calculating our estimates, we consider the same matrix of reporting probabilities \( P \) and the same basis functions used to generate data. We also assume that \( \Sigma_c = \sigma^2_c I \). As described in Section 3.4, we approximate \( H_i \) in transformer \( i = 1, \ldots, I \) by simulating \( B = 100,000 \) independent data sets just once and tabling the results.

The estimates of \( \alpha_1, \alpha_2, \) and \( \alpha_3 \) are summarized in Figures 3 and 4, which show the first and third quartiles of the 200 estimates for each of Cases 1 and 2, along with the “true” \( \alpha_c \)’s. We see that estimates of the \( \alpha_c \)’s are best when the true \( \alpha_c \)’s are of the same scale (Figure 3). As expected, estimates of the \( \alpha_c \)’s are less variable when they are based on data with more transformers (bottom two rows of Figures 3 and 4). We always obtain very good estimates of the total electric load for each transformer through the weighted sum \( \hat{M}_1 \alpha_1(t) + \hat{M}_2 \alpha_2(t) + \hat{M}_3 \alpha_3(t), \ i = 1, \ldots, 5 \). We see this even when \( \alpha_1 \) and \( \alpha_2 \) are smaller than \( \alpha_3 \), and the estimates are severely biased (Figures 4 and 5).

Figure 6 presents boxplots of the bias in estimating the true number of consumers in each class in the first three transformers. We can see that usually estimates are within 1 or 2 units of the true value. On closer examination of results (not shown here), we find that the most common estimate of \( M \) is either at the reported number or at the true value of \( M \) or at some number in between.

For variance component estimation, the regression error variance is very well-estimated in all simulation scenarios, as seen in Table 2. However, the variances of the random effects were not so well estimated. Indeed, many estimates were equal to 0, as we can see in Table 3. The pattern of occurrences of zero estimates is what one would expect with method of moments type estimators: when variances are small and/or sample size is small, many variance estimates are negative (and so set equal to 0). In Case 1, when \( \alpha_1 \) and \( \alpha_2 \) are of the same scale, \( \sigma^2_{\gamma,1} = 0.09, \sigma^2_{\gamma,2} = 0.007, \) and \( \sigma^2_{\gamma,3} = 0.025, \) so \( \sigma^2_{\gamma,2} \) is particularly challenging to estimate. In Case 2, when \( \alpha_1 \) is much smaller than \( \alpha_2, \sigma^2_{\gamma,1} = 0.089, \sigma^2_{\gamma,2} = 0.019 \) and \( \sigma^2_{\gamma,3} = 0.73, \) so \( \sigma^2_{\gamma,3} \) is the easiest to estimate.

To investigate the effect of increasing the number of transformers as well as increasing the number of replicates (number of days), we conduct further simulations and present the results in the Supplementary Material. To speed computation, we consider only two classes, but we extend our studies to consider four cases:

Case A: The two functions \( \alpha_1 \) and \( \alpha_2 \) are of the same scale, and the \( M_1 \)'s and \( M_2 \)'s are balanced.

Case B: The two functions \( \alpha_1 \) and \( \alpha_2 \) are of the same scale, and the \( M_1 \)'s are much bigger than the \( M_2 \)'s.

Case C: The function \( \alpha_1 \) is of a much smaller scale than the function \( \alpha_2 \) and the \( M_1 \)'s and \( M_2 \)'s are balanced.

Case D: The function \( \alpha_1 \) is of a much smaller scale than the function \( \alpha_2 \) and the \( M_1 \)'s are much bigger than the \( M_2 \)'s.

For each of the four cases, we carry out six simulation studies: (a) with 5 transformers and either no replicates or 5, 30 or 100 replicates and (b) with 50 transformers and either no replicates or 5 replicates. As expected, the variability of the estimated typologies is reduced by increasing the number of transformers and/or the number of replicates. In addition, the bias in typology estimation is reduced by increasing the number of transformers. Perhaps surprisingly, this bias is not reduced by increasing the number of replicates. Thus, in cases where estimates have a large bias and a decreased variability, the estimated typologies concentrate around the wrong curve. Even with the increased number of replicates and/or number of transformers, when \( \alpha_1 \) is much smaller than \( \alpha_2 \) (Cases C and D), many estimates of \( \sigma^2_{\gamma,1} \) were zero. Estimation of \( \sigma^2 \) was good throughout. In general, we found that estimation of \( M_1 \) was as described above: the estimates tended to be shifted from the reported number of consumers of Class 1 to the true number. The variability of the estimates of the \( M \)'s decreased with the number of replicates but not with the number of transformers.

7 | CONCLUSIONS

In this paper, we proposed a generalization of the work of Dias et al. (2009) on estimating mean curves when the available sample consists of aggregated functional data. The main novelty of this work is to incorporate a randomness in the counts for class membership. This flexibility allows the analysis of the data even when there is some misreporting in the number of consumers of each type. We also use random effects to model the within transformer correlation structure.
To study the properties of our method, we analyzed artificial data sets exploring different scenarios, and we also analyzed a real data set. The artificial data allowed us to explore the influence of increasing the number of replications and the number of transformers. In the data example, comparing the observed aggregated curves with the weighted sum of the estimated typical ones, we see that the proposed model provides reasonable estimates of the mean aggregated curve.

We do believe that our work can be generalized relaxing two crucial assumptions made in this study:

1. The total number of consumers served by transformer \(i\), \(N_i\), is equal to \(\sum_{c=1}^{C} R_{ci}\), which is also equal to \(\sum_{c=1}^{C} M_{ci}\); and
2. There is a known matrix of reporting probabilities \(P\), not depending on the transformer, with \(P(c, r) = P[R_{ci} = r|M_{ci} = c]\), \(r, c = 1, \ldots , C\).

Other points to be explored are the facts that energy consumption is subject to diurnal, weekly and seasonal variability and that energy use on weekdays is typically different from that on weekends and holidays. At this point, we do not have enough data to estimate these sources of variability. However, if we were given quarter-hour measurements each day for 10 years, for example, we could include an additional fixed effect to account for seasonal differences and another fixed effect plus an indicator variable in the modeling of weekend/holiday energy consumption. To do so, we would still allow \(t\) to be measured in hours, but with \(t\) ranging over 10 years. We let \([t]\) denote the “hour of the day”, equal to \(t\) modulo 24. We could then model \(Y_i(t)\), the observed aggregated data for transformer \(i\) at time \(t\) as

\[
Y_i(t) = \sum_{j=1}^{N_i} W(t_j) + e(i, t),
\]

with

\[
W(t)_{ci} = \alpha_c([r]) + \alpha_{ch}([r]) \times I[\text{weekend/holiday}] + S_c(t) + \alpha_{ci}^t(t),
\]

where \(\alpha_c\) and \(\alpha_{ch}\) are the nonrandom typical usage curves in class \(c\), \(S_c\) is a flexibly modeled periodic function with period one year, to accommodate seasonality, and \(\alpha_{ci}^t\) is the consumer-specific random perturbation, which could be modeled as periodic with period of 1 year, if desired. Note that \(\alpha_c\) and \(\alpha_{ch}\) do not depend on day of year but should be modeled to ensure continuity in expected energy consumption.

The realistic estimation of consumption patterns is a very complex problem that can only be solved through the analysis of simpler pieces. This paper is a first attempt in this direction, incorporating variability in the number of consumers.

**APPENDIX**

**Proof. (Proof of Theorem 1)**

Write \(P\{R = r|M = m\}\) as

\[
\sum_{x,c:|x| = m} \{P\{X_1 = (x_{11}, \ldots , x_{1C})\} \times P\{X_2 = (x_{22}, \ldots , x_{2C})\} \times \cdots \times P\{X_C = (x_{C1}, \ldots , x_{CC})\}\} = \sum_{x,c:|x| = m} \left\{ \frac{m_1!}{x_{11}! \cdots x_{1C}!} P(1, 1)^{x_{11}} \cdots P(1, C)^{x_{1C}} \times \cdots \times \frac{m_C!}{x_{C1}! \cdots x_{CC}!} P(C, 1)^{x_{C1}} \cdots P(C, C)^{x_{CC}} \right\}. \tag{A.1}
\]

Rearranging terms yields that the probability is equal to

\[
\frac{\prod_{c=1}^{C} m_c!}{\prod_{c=1}^{C} r_c!} \sum_{x,c:|x| = m} \left\{ \frac{r_1!}{x_{11}! \cdots x_{1C}!} P(1, 1)^{x_{11}} P(2, 1)^{x_{22}} \cdots P(1, C)^{x_{1C}} \right\} \times \left\{ \frac{r_C!}{x_{C1}! \cdots x_{CC}!} P(C, 1)^{x_{C1}} \cdots P(C, C)^{x_{CC}} \right\} \times \left( \sum_{c=1}^{C} P(c, j) \right)^{r_j} - \left( \sum_{c=1}^{C} P(c, j) \right)^{r_j}
\]

\[
= \frac{\prod_{c=1}^{C} m_c!}{\prod_{c=1}^{C} r_c!} \times \prod_{c=1}^{C} m_c! \times \sum_{x,c:|x| = m} P\{X_j = (x_{1j}, \ldots , x_{Cj})\}
\]

\[
= \left( \sum_{c=1}^{C} P(c, j) \right)^{r_j} \times \prod_{c=1}^{C} m_c! \times H(m_1, \ldots , m_C).
\]

\(\square\)
SUPPLEMENTARY MATERIAL

The file contains the supplementary material for this manuscript. In Section 1 we present a simple example of how to approximate $P(R = r|\mathbf{M} = \mathbf{m})$ for fixed $r$ via one simulation study. Section 2 includes the results of further simulation studies for different numbers of transformers and replicates, when we only consider two consumer classes.

ACKNOWLEDGEMENTS

This work was partially funded by CNPq grants 442012/2014-4, 3025981/2014-6, 476476/2013-5 and 302956/2013-1, FAPESP 2014/26419-0, 2014/26414-9 and 2013/07375-0 (CEMAI-CEPID), CAPES, CBIE-ELAP and NSERC grant 7969. The authors are grateful to CPFL for providing the data and the University of British Columbia and the University of Washington for their hospitality.

REFERENCES

Bodin, P., Villemoes, L.F., & Wahlberg, B. (2000). Selection of best orthonormal rational basis. SIAM Journal on Control and Optimization, 38(4), 995–1032 (electronic).

Cardoso, J.-F. (1998). Blind signal separation: Statistical principles. Proceedings of the IEEE, Special Issue on Blind Identification and Estimation, 9(10), 2009–2025.

Choi, S., Cichocki, A., Park, H.-M., & Lee, S.-Y. (2005). Blind source separation and independent component analysis: A review. Neural Information Processing - Letters and Reviews, 6(1), 1–57.

Comon, P., & Jutten, C. (2010). Handbook of Blind Source Separation: Independent Component Analysis and Applications (1st ed.). New York: Academic Press.

De Vore, R., Petrova, G., & Temlyakov, V. (2003). Best basis selection for approximation in $L^p$. Foundations of Computational Mathematics, 3(2), 161–185.

Dias, R. (1998). Density estimation via hybrid splines. Journal of Statistical Computation and Simulation, 60, 277–294.

Dias, R., Garcia, N.L., & Martarelli, A. (2009). Non-parametric estimation for aggregated functional data for electric load monitoring. Environmetrics, 20(2), 111–130.

Dias, R., Garcia, N.L., & Schmidt, A.M. (2013). A hierarchical model for aggregated functional data. Technometrics, 55, 321–334.

Luo, Z., & Wahba, G. (1997). Hybrid adaptive splines. Journal of the American Statistical Association, 92, 107–116.

Ramsay, J.O., & Silverman, B.W. (2005). Functional Data Analysis, New York: Springer.

Tseng, P. (2001). Convergence of a block coordinate descent method for nondifferentiable minimization. Journal of Optimization Theory and Applications, 109(3), 475–494.

Wahba, G. (1983). Bayesian “confidence intervals” for the cross-validated smoothing spline. Journal of the Royal Statistical Society Series B, 45, 133–150.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Lenzi A, de Souza CPE, Dias R, Garcia NL, Heckman NE. Analysis of aggregated functional data from mixed populations with application to energy consumption, Environmetrics. 2017;28:e2414. https://doi.org/10.1002/env.2414