Simultaneous confidence bands in curve prediction applied to load curves

Jean-Marc Azaïs, Sophie Bercu, Jean-Claude Fort, Agnes Lagnoux, Pierre Lé

Simultaneous confidence bands in curve prediction applied to load curves

J.M. Azaïs[1], S. Bercu[2], J.C. Fort[1], A. Lagnoux[1]and P. Lé[2].

[1]*Institut de Mathématiques de Toulouse, Université de Toulouse, Toulouse, France.*

[2]*EDF Recherche & Développement, Département ICAME, Clamart, France.*

Abstract

In numerous contexts, one has to forecast a full curve from some explanatory variables. For that purpose, one aims at deriving simultaneous confidence bands. In this article theoretical and numerical results on the maximum of Gaussian sequences are used to construct those simultaneous confidence bands.

Keywords : Gaussian vectors, maximum of Gaussian sequences, simultaneous confidence bands, thresholding, load curve.

1 Introduction

In curve prediction, one is generally interested in deriving simultaneous confidence bands: regions in which the entire curve lies with some given probability. Simultaneous confidence bands have been widely studied in the literature and the topic has known an incredible expansion since the first articles published in the early fifties. The first studies have been conducted in the context of linear models by Tukey [26] and [24]. In the last paper, the construction of the simultaneous confidence bands relies on the Fisher distribution and the confidence region is an ellipsoïd. The principle of Scheffé's method has been also generalized to non linear regression and usually relies on Bootstrap and asymptotic results (see e.g. the articles of Claeskens et al. [4], Hall [10] and the references therein). These techniques have been also extended and adapted for various problems (see e.g. estimation of distribution [3,11], probability functions [5], elements of the spectral density matrix of autoregressive processes [23], approximation of an integral by Monte Carlo method [14]...). Here we introduce a new technique based on the maximum of Gaussian processes whose distribution is estimated by a MCQMC algorithm proposed by Alan Genz ([8,9]) on one hand and Sidak's inequality on the other hand, which seems to be new.

We implement this technique in the numerical context of load curve prediction: power producers like EDF, the electrical French group, use the information contained in the load curves of their customers to plan electricity production and to offer them an appropriate tarification. Since this

information is not available for all the clients, EDF needs to estimate these load curves from a learning sample. To appreciate the accuracy of these estimations we are interested in deriving precise simultaneous confidence bands. For that purpose, we construct a model (involving three main parameters, one of which is the weekly cycle) to represent precisely the load curves from explanatory variables. After a linear regression of the curves on the explanatory variables, we use the proposed technique. Obviously, this technique may apply to any context in which one need to estimate the evolution of some observed quantity during the time by giving simultaneous confidence bands. As a consequence, the construction of such bands represents a great challenge in numerous and various applied domains and particularly in industry (e.g. signal detection by Hall et al. [12] or Macskassy et al. [18]) or biology (e.g. Zouhourian et al. [27])...

The article is organized as follows. Section 2 describes the general framework and the procedure to construct the estimates and to derive the simultaneous confidence bands. Section 3 is dedicated to the computation of the maximum of the absolute value of a Gaussian series. Section 4 presents the framework of the EDF context, while numerical results are given in Section 5. The article ends with conclusion and prospects in Section 6.

2 Estimation of curves in high dimension using explanatory variables

In this section, we consider a family of curves $R_k(x) \in \mathbb{R}$ associated to the individual k ($k = 1, \ldots, K$) and for a generic time x, integer in $\mathcal{X} = 1, \ldots, \mathbf{x}$. In the EDF context detailed in Section 4, it can be for example

- the entire (annual) load curve $(\mathrm{LC}_k(t))_t$ of the customer k, $k = 1, \ldots, K$, over time $t = 1, \ldots, 17520$. t represents any half-hour in the year.
- the centered and standardized week load curve $(S_k(h))_h$ of this customer, over time $h = 1, \ldots, 336$. h represents any half-hour in the week.
- the mean of every week load curve $(M_k(w))_w$ of this customer, over time $w = 1, \ldots, 52$. w represents any week in the year.

2.1 Linear regression

We assume that each curve $R_k(.)$ belongs, up to an error term, to some linear functional space i.e.

$$R_k(x) = \sum_{j \in \mathcal{J}} a_k^j \phi_j(x) + \epsilon_k(x) \qquad (k = 1, \ldots, K) \tag{1}$$

where $\epsilon_k(x)$ is an extra term assumed to be a Gaussian white noise and $(\phi_j)_{j\in\mathcal{J}}$ is an orthonormal basis. We denote $J = \text{card}(\mathcal{J})$.

We do not observe directly the coefficients a_k^j but

$$\tilde{a}_k^j = a_k^j + \beta_k^j, \quad k = 1, \ldots, K, \quad j = 1, \ldots, J, \tag{2}$$

where (β_k^j) is an extra error term.

Moreover we suppose that to each individual are associated P explanatory variables and that for each j a linear regression model is used:

$$a_k^j = \sum_{p=1}^{P} \alpha_p^j V_{k,p} + \eta_k^j = V(k)\alpha^j + \eta_k^j, \quad k = 1, \ldots, K, \tag{3}$$

where $V = (V_{k,p})_{(k=1\ldots K, p=1\ldots P)}$ is the matrix of the regression model and $V(k)$ the k-th row of V. Thus we have $J = \text{card}(\mathcal{J})$ linear models of the form (3) that can be rewritten as

$$a^j = V\alpha^j + \eta^j, \text{ for } j = 1, \ldots, J. \tag{4}$$

Eventually,

$$\tilde{a}_k^j = V(k)\alpha_k^j + \eta_k^j + \beta_k^j, \quad k = 1, \ldots, K, \quad j = 1, \ldots, J. \tag{5}$$

Remark 1

- *In most cases, the model will contain a constant: $V_{k,1} \equiv 1$.*
- *We assume that the errors of the J models, defined by (3), are independent with different variances. Since $\epsilon_k(x)$ is a Gaussian white noise, the errors of the J models, defined by (5), are also independent with different variances σ_j^2.*
- *Linear regression is only an example; more complicated models, as nonlinear regression, can be studied at the cost of the use of some asymptotic results.*

Let $\hat{\alpha}_p^j$ the least square estimator of α_p^j, $j = 1, \ldots, J$ and $p = 1, \ldots, P$. The estimation of a_k^j is then given by:

$$\hat{a}_k^j = \sum_{p=1}^{P} \hat{\alpha}_p^j V_{k,p} = V(k)\hat{\alpha}^j. \tag{6}$$

with $\text{Var}((\hat{a}_k^j)_{k=1,\ldots,K}) = \sigma_j^2 V(V'V)^{-1}V'$ and obviously an estimation or a prediction of the response $R_k(x)$ is

$$\hat{R}_k(x) = \sum_{j\in\mathcal{J}} \hat{a}_k^j \phi_j(x), \tag{7}$$

with

$$\mathrm{Var}\left(\hat{R}_k(x)\right) = \sum_{j\in\mathcal{J}} \phi_j^2(x)\sigma_j^2 \left(V(V'V)^{-1}V'\right)_{k,k}, \tag{8}$$

where $\left(V(V'V)^{-1}V'\right)_{k,k}$ is the (k,k)th element of $\left(V(V'V)^{-1}V'\right)$. Note that $\hat{R}_k(x)$ is an unbiased estimator of $R_k(x)$.

2.2 Non simultaneous confidence bands

Each estimation $\hat{\sigma}_j^2$ of the variance σ_j^2 will be considered as exact because of the size of the samples. The construction of a non simultaneous confidence interval for $R_k(x)$ relies on the following fact

$$Z_k(x) := \frac{\hat{R}_k(x) - R_k(x)}{\sqrt{\sum_{j=1}^J \phi_j^2(x)\hat{\sigma}_j^2 \left(V(V'V)^{-1}V'\right)_{k,k}}} \sim \mathcal{N}(0,1),$$

thanks to the CLT. As a consequence, if z_α is the α quantile of the standard Gaussian distribution, we have for each (x,k)

$$|Z_k(x)| < z_{1-\alpha/2} \Leftrightarrow R_k(x) \in \hat{R}_k(x) \pm z_{1-\alpha/2}\sqrt{\sum_{j=1}^J \phi_j^2(x)\hat{\sigma}_j^2 \left(V(V'V)^{-1}V'\right)_{k,k}}.$$

with probability $1-\alpha$ and we get confidence intervals for $R_k(x)$ as x varies which are not simultaneous.

2.3 Simultaneous confidence bands

In this section, our aim is to construct simultaneous confidence bands for each individual k i.e. to determine for the given $k \in K$ a fractile function $F_{1-\alpha}^k(x)$ such as

$$\mathbb{P}\left\{\forall x \in \mathcal{X} : R_k(x) \in \left[\hat{R}_k(x) \pm F_{1-\alpha}^k(x)\right]\right\} = 1-\alpha.$$

Different techniques to build simultaneous confidence bands are encountered in the literature as mentioned in the introduction; e.g. Scheffé method which is not tractable in our framework since we do not require simultaneous confidence bands both in k and x. We propose in this section a new technique based on the supremum of Gaussian random variables, see Section 3.

Under the previous hypothesis and with the same notation, for all k, remind that $(Z_k(x))_x$ defines a Gaussian centered and standardized sequence with covariance given by:

$$\mathrm{Cov}\left(Z_k(x), Z_k(x')\right) = \frac{\sum_{j=1}^J \phi_j(x)\phi_j(x')\hat{\sigma}_j^2}{\sqrt{\sum_{j=1}^J \phi_j^2(x)\hat{\sigma}_j^2 \sum_{j=1}^J \phi_j^2(x')\hat{\sigma}_j^2}}$$

So if we are able to derive the value $S_{1-\alpha}$ such that

$$\mathbb{P}\left(\sup_{x=1,\ldots,\mathbf{x}} |Z_k(x)| \leq S_{1-\alpha}\right) = 1-\alpha,$$

we will get simultaneous confidence bands as required. The distribution of the supremum of a Gaussian process is of great interest but very few exact theoretical results can be found in the literature.

3 The computation of the maximum of the absolute value of a Gaussian series

In the following $Z_1, \ldots, Z_n$ consist of n observations of a centered Gaussian series which is not necessarily stationary but has a constant variance. Without loss of generality, we assume that this variance is 1. We are interested in the distribution of the variables

$$M_n^* := \sup_{i=1,\ldots,n} |Z_i| \quad \text{or} \quad M_n := \sup_{i=1,\ldots,n} Z_i.$$

3.1 Classical inequalities

In this section we review classical bounds. The first one developed by Knowles in his 1987 Stanford dissertation is based on the work of Naiman [22] and Hotelling [13] on the volume of tubes. Define the correlation between X_{j-1} and X_j by ρ_j and the length of the sequence $(X_1, \ldots, X_n)$ by $L := \sum_{j=2}^{n} \arccos(\rho_j)$. The length-based bound on $\mathbb{P}\{M_n \geq y\}$ is then

$$\mathbb{P}\{M_n \geq y\} \leq 1 - \Phi(y) + e^{-y^2/2}\frac{L}{2\pi} \tag{9}$$

(where Φ is the distribution function of a standard Gaussian variable) and usually beats the crude Bonferroni bound (Miller [19]) which is known to be rough:

$$\mathbb{P}\{M_n \geq y\} \leq n(1-\Phi(y)). \tag{10}$$

The following bound (called W in the sequel) is a natural improvement of the length-based one and have been developed by Efron [7]. It still involves the correlation between two successive variables:

$$\mathbb{P}\{M_n \geq y\} \leq 1 - \Phi(y) + \phi(y)\sum_{j=1}^{n} \frac{\Phi(yL_j/2) - 1/2}{y/2} \tag{11}$$

where $L_j = \arccos(\rho_j)$ and ϕ is the density function of a standard Gaussian variable.

Finally, Sidak's inequality [25] (called Sidak in the sequel) applies in any context (in particular when the correlation structure is unknown) and shows that under weak conditions the independent case is the worst case in the sense that

$$\mathbb{P}\{M_n^* \leq y\} \geq (2\Phi(y) - 1)^n \tag{12}$$

So using the critical value for an independent sequence will always lead to a conservative test.

For very large n one must use an efficient version of the quantile function of the normal distribution. This is in general easy and it is the case for example in Matlab with the function norminv.

3.2 $n \leq$ (say) 500 or 1000: almost exact calculation by MC-QMC

In such a case the MCQMC Matlab program QSIMVNV written by Genz [8] allows the direct calculation of Gaussian probability over hyper-rectangle for dimension up to 500-1000. It consists of transforming the integral into an integral over the hyper-cube $[0,1]^n$ and then using quasi-Monte-Carlo (QMC) integration with lattice rule. In a final step the procedure is randomized using a Monte-Carlo quasi-Monte-Carlo (MCQMC) method. See Genz [8] or Azaïs and Genz [1] for more details. The routine QSIMVNV provides an estimation of its numerical error.

3.3 Asymptotic band from extreme value

It is well known [17] that under very weak assumptions, for stationary sequences,

$$\mathbb{P}\big(a_n(M_n - b_n) \leq y\big) \to \exp(-\exp(-y)) \quad \text{as} \quad n \to +\infty \tag{13}$$

$$\text{with} \quad \left\{ \begin{array}{l} a_n := (2\log n)^{\frac{1}{2}} \\ b_n := (2\log n)^{\frac{1}{2}} - \frac{1}{2}(2\log n)^{-\frac{1}{2}} \left(\log\log n + \log 4\pi\right). \end{array} \right.$$

Using symmetry, a bound can be deduced for M_n^* using the $\alpha/2$ bound for M_n. These results can be easily extended to non-stationary series using the tools of Azaïs and Mercadier [2] (In fact this paper considers the more complicated case of random processes). The quality of this approximation is considered in Section 3.4.

Note that in the case where n is very large, the considered levels are very high and some small deviation from normality can heavily affects the results.

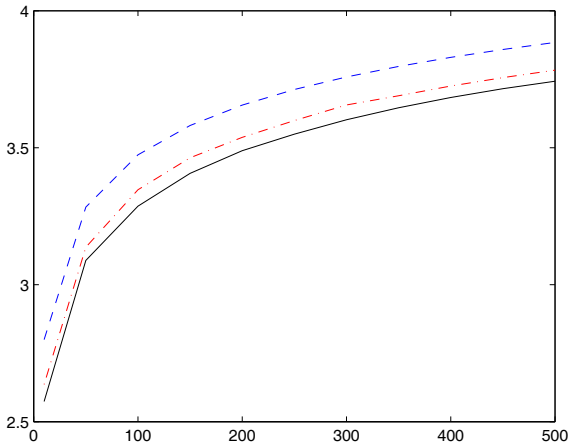

Figure 1: For the AR(1) series with correlation 0.9, comparison of the 5% critical values computed with Sidak, W and the exact algorithm (from top to bottom) as a function of the length of the series.

Remark also that the crude approximation $\sqrt{2 \log n}$ is often used in practice.

3.4 Numerical experiment

The first application considers the comparison of the exact calculation of Genz with Sidak and W for an AR(1) series with parameter 0.9. Here the correlation between two successive variables can easily be determined. Figure 1 compares the 5% critical values as a function of n. It shows clearly that the exact calculation provides an improvement for small values and that improvement decreases with the size. Moreover W is more accurate than Sidak. Table 1 gives three examples.

Table 1: Numerical experiment

value of n	10	500	900
Exact bound	2.56	3.74	3.90
Sidak	2.80	3.88	4.02
difference	0.24	0.14	0.12
W	2.64	3.79	3.93
difference	0.08	0.05	0.03

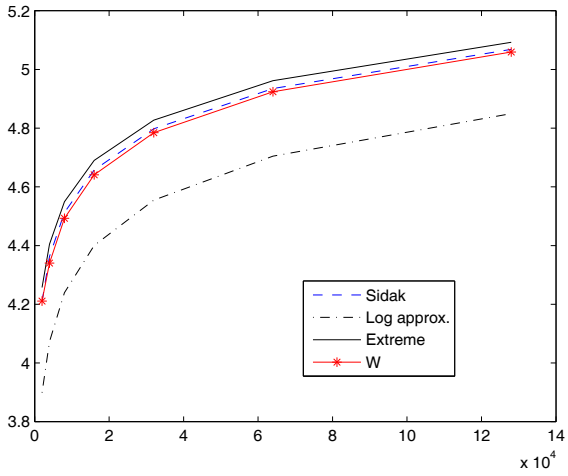

Figure 2: For the 5% critical value, comparison of Extreme, Sidak, W (with $\rho = 0.8$) and $\sqrt{2\log(n)}$ (from top to bottom) for large values of n

The interpretation of Figure 2 is not easy since the true result is not known and may depend on the particular covariances of the processes. Nevertheless we believe that this dependence is small because the extreme result (Extreme) shows it is true asymptotically. The main points are

- Sidak is better than Extreme since it controls the level and is smaller than Extreme.
- There is some coherence between Extreme, W and Sidak and we believe that for most processes the true result is close to these three. As a consequence the $\sqrt{2\log n}$ approximation seems very crude.
- Our conclusion is that for very large n, Sidak and W are the safer choices and give very similar results. In the application section we have chosen Sidak for simplicity.

4 Application to the EDF load curves

4.1 General settings

As mentioned in the Introduction, we apply the technique presented in the article in the context of EDF load curves prediction. A load curve is a chart showing the amount of electrical energy that a customer uses over the course of time. Power producers like EDF, the electrical French group, use this information to plan how much electricity they need to make available

at a given time. But this load curve is only available for customers with automated meter reading. For the others, it must be estimated by EDF in order to know their profitabilities and offer them an appropriate setting of prices. In this EDF context, different approaches have been already studied: a non parametrical approach developed by Misiti et al. [20], a wavelet analysis by Misiti et al. [21].

To achieve this prediction problem, we used a data set of 917 customers containing for each of them their annual load curve valued every half-hour (which corresponds to 17520 points) and some explanatory variables (17 exactly). These explanatory variables are of different nature: qualitative (activity sector, geographical localization...) and quantitative (total annual consumption, ratio of consumption summer/winter, ratio rush/slack hours...). Customers are in general firms and will be denoted as "individuals" in the sequel. We split data into a learning sample of 832 individuals and a test sample of 85 individuals. These sizes have been chosen heuristically.

For the sake of simplicity, we remove the index k of individual in the sequel.

4.2 The decomposed model

In this paragraph, we describe a model, called "decomposed" model, that takes into account the weekly cycles and lies on three main quantities:

- the centered and standardized week load curve $(S(h))_h$ of the given customer;
- the mean of this week load curve $(M(w))_w$;
- the standard deviation of this week load curve $(\sigma(w))_w$.

More precisely, the decomposed model consists first in constructing a typical week for each of the individuals: for a given individual k and a half-hour h of the week, the value of the typical week at time h is simply the median of all the values of the load curve of the individual k at the same half-hour h of the week. The obtained curve is centered and standardized and the reconstruction is then done by multiplying the typical week by a standard deviation and adding it a mean level. We explain that procedure more precisely in the rest of the section. "Time" is here measured in half an hour.

First of all, let us introduce some notation (recall that k is fixed and thus removed for clarity):

- $\mathrm{LC}(t)$ the entire load curve at time $t = 1, \ldots, T = 17520$,

- $w(t) = 1, \ldots, 53$ and $h(t) = 1, \ldots, 336$ are respectively the index of the week and the intra week time associated to Time t. Note that in year 2002, the first and last weeks, with index 1 and 53, are incomplete and contain respectively 6 and 2 days. For these weeks $h(t)$ varies from 1 to 288 and from 1 to 96 respectively.
- $C^{Tot} := \sum_{t=1}^{17520} \mathrm{LC}(t)$ the total consumption which is known.

Each load curve is decomposed in the following manner

- To work on the rescaled profiles of the annual load curves (without dimension), it is normalized by dividing by the total annual consumption. We denote by $RS^1(t)$ the remaining signal.
- The mean of every week

$$M(w) = \frac{1}{336} \sum_{w(t)=w} RS^1(t)$$

(with obvious modification for the first and last weeks) is subtracted

$$RS^2(t) = RS^1(t) - M(w(t))$$

- Each week is standardized. We compute first the empirical variance of week w

$$\sigma^2(w) = \frac{1}{336} \sum_{w(t)=w} (RS^2(t))^2$$

and we define

$$RS^3(t) = \frac{1}{\sigma(w(t))} RS^2(t)$$

- A typical centered and standardized week is computed

$$S(h) = \mathrm{median}\{RS^3(t), \quad h(t) = h\}$$

In fact a small modification of this formula is done to consider bank holidays as Sundays (see [15] for more details).

Eventually the reconstruction is performed in the following way

$$\mathrm{LC}(t) \quad = \quad [S(h(t))\sigma(w(t)) + M(w(t))]\, C^{Tot} + \nu(t) \tag{14}$$

for $t = 1, \ldots, 17520$.

Formula (14) must be considered as a decomposition of each observed curve

LC and not as an estimation in the statistical sense. We assume only that the resulting noise ν, which is due to the replacement of the week by the median, is Gaussian and with constant variance and decaying covariances. In practice, this assumption is approximately satisfied since $\nu(t)$ is in general small. This decomposition has shown to be efficient in previous studies at EDF see e.g. [15].

Remark 2 *For the functions $R_k = S(h)$, $\sigma^2(w)$ and $M(w)$, we use a preliminary compression on the Fourier basis. Fourier basis has been chosen because of the periodicities that are encountered in the problem. We choose the coefficients, globally for all individuals, by a hard thresholding procedure to get model (1). Since this technique is not the purpose of the current paper, the reader should refer to [6] for details and other references therein.*

4.3 Estimation and simultaneous confidence bands of the load curve

- To have a comparison scale (which is not the main point of this article), we also consider the model with no decomposition called the global model. In this model, the method is based on
  - a compression of $(\mathrm{LC}(t))_t$:

$$\mathrm{LC}(t) = \sum_{j=1}^{J} a^j \phi_j(t) + \epsilon(t) =: \overline{\mathrm{LC}}(t) + \epsilon(t)$$

    First, $\overline{\mathrm{LC}}(t)$ is estimated by $\widehat{\mathrm{LC}}(t)$ following the procedure detailed in Section 2.1. Second, we estimate the variance of $\epsilon(t)$ by

$$\sigma_\epsilon^2 = \frac{1}{K} \sum_{k=1}^{K} \left\{ \frac{1}{T-1} \sum_{t=1}^{T} \left[\mathrm{LC}_k(t) - \overline{LC}_k(t) \right]^2 \right\}.$$

    We suppose moreover this estimation exact because it is based on 832×17520 observations.

  - the fact that

$$\frac{\widehat{\mathrm{LC}}(t) - \overline{\mathrm{LC}}(t)}{\mathrm{Var}\left(\widehat{\mathrm{LC}}(t)\right)} \sim \mathcal{N}(0,1),$$

    where $\mathrm{Var}\left(\widehat{\mathrm{LC}}(t)\right)$ is the expression (8).

As a consequence for any individual,

$$\mathrm{LC}(t) - \widehat{\mathrm{LC}}(t) = \left[\overline{\mathrm{LC}}(t) - \widehat{\mathrm{LC}}(t)\right] + \epsilon(t)$$

has the variance $\mathrm{Var}\left(\widehat{\mathrm{LC}}(t)\right) + \sigma_\epsilon^2$. And we are able to construct confidence bands:

$$\left[\widehat{\mathrm{LC}}(t) \pm S_{1-\alpha}\sqrt{\sum_{j\in\mathcal{J}} \phi_j^2(t)\sigma_j^2 \left(V(V'V)^{-1}V'\right)_{k,k} + \sigma_\epsilon^2}\right]$$

where $S_{1-\alpha}$ is simply 1.96 in the case of non simultaneous confidence intervals and is given by Sidak (Section 3.1, equation (12)) in the case of simultaneous confidence bands.

- For the decomposed model, we use the decomposition (14) and set

$$\overline{\mathrm{LC}}(t) = \left[\overline{S}(h(t))\overline{\sigma}(w(t)) + \overline{M}(w(t))\right] C^{Tot} + \epsilon(t).$$

Now we apply the method of Section 3.2 (MCQMC/Genz method for $n = \mathbf{x} = 336, 53$ and 53 respectively) to every component, namely

$$\overline{S}, \overline{\sigma}, \overline{M}$$

with level $\alpha/3$, to derive estimation, lower and upper bounds for these quantities. Recall that C^{Tot} is known. We then plug these bounds into decomposition (14)to get upper and lower bounds of $\overline{\mathrm{LC}}(t)$:

$$\overline{\mathrm{LC}}(t) \in \left[\widehat{\mathrm{LC}}(t) \pm f_{1-\alpha}(t)\right]$$

with probability greater than $1-\alpha$. Then we estimate the variance of

$$\epsilon(t) = \mathrm{LC}(t) - \overline{\mathrm{LC}}(t)$$

and we use Sidak to get

$$|\epsilon(t)| = |\overline{\mathrm{LC}}(t) - \mathrm{LC}(t)| \leq \sigma_\epsilon B_{\mathrm{Sidak}},$$

where B_{Sidak} is the bound obtained following procedure of Section 3.1. For any individual, we construct a simultaneous confidence region of the form

$$|\widehat{\mathrm{LC}}(t) - \mathrm{LC}(t)| \leq f_{1-\alpha}(t) + \sigma_\epsilon B_{\mathrm{Sidak}}.$$

Prediction quality and covering ratios

To evaluate the accuracy of the predictions, we define an error criterion weighted by the energy cost w. For each individual, we compute a quadratic relative error normalized by the typical price curve of energy:

$$\mathrm{Err} = \frac{\sum_t \left| \mathrm{w}(t) \Big(\mathrm{LC}_{\mathrm{predicted}}(t) - \mathrm{LC}_{\mathrm{real}}(t) \Big) \right|^2}{\sum_t \left| \mathrm{w}(t)\, \mathrm{LC}_{\mathrm{real}}(t) \right|^2}$$

Err is the error done for an individual whose real load curve is $\mathrm{LC}_{\mathrm{real}}$, predicted curve is $\mathrm{LC}_{\mathrm{predicted}}$ and w represents a possible weighting (e.g. by prices).

To evaluate the prediction performance on the test sample, we define two covering rates: for the simultaneous confidence bands (SCB),

$$\mathrm{SCR} := \frac{\#\{k : \text{the entire curve is in the SCB}\}}{\#\{k\}};$$

and for the non simultaneous confidence bands (NSCB),

$$\mathrm{NSCR} := \frac{\#\{(k,t) : \mathrm{LC}_k(t) \in \mathrm{NSCB}\}}{\#\{(k,t)\}}.$$

Moreover, for $\rho \in [0,1]$, we define $\mathrm{SCR}(\rho)$ as the ratio of load curves having at least a percentage ρ of their points in the simultaneous confidence bands. Obviously, with these definitions $\mathrm{SCR} = \mathrm{SCR}(1)$.

5 Numerical results

5.1 Learning sample

In this section, we present briefly a few results.

Decomposed model
In the study of typical weeks S, hard thresholding suggests us to achieve 99% of the variance which corresponds to keep the 105 best frequencies (i.e. 210 best coefficients). We plot in Figure 3 the typical week of individual 11, its estimation and confidence bands (simultaneous and non simultaneous). One can find the same type of figure for other individuals in [16].

We determine estimate and confidence bands for the other components in the same way. The construction of the estimates of the entire curve is then done in a natural way using the equation of reconstruction (14) given in Section 4.2. Concerning the confidence bands, the operation is more tricky.

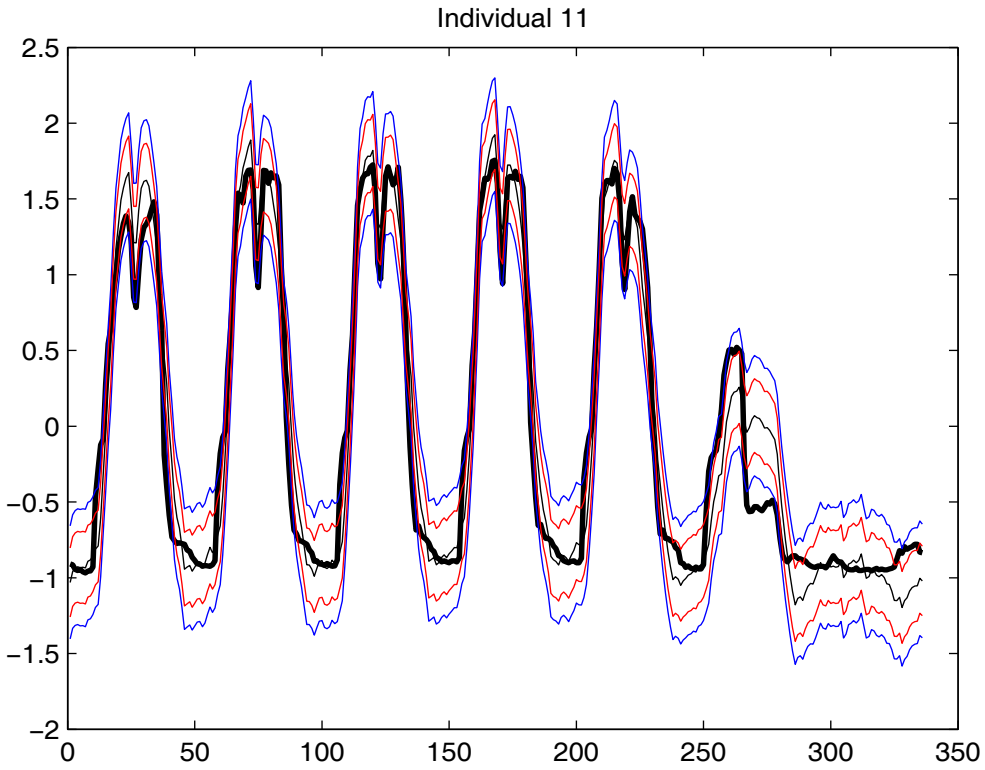

Figure 3: Estimation of the typical week S of individual 11 and associated confidence bands : the typical week is represented by the thick line, its estimation is the central curve and on both sides of this curve the confidence bands (the closest are the non simultaneous bands and the largest are the simultaneous bands)

The three quantities being non negative, we multiply the lower and upper confidence bounds for each of the components studied following the same equation (14). Nevertheless, in that way the level is not conserved; thus in order to have a global confidence level of $1-\alpha$, we need to construct for each of the components confidence bands with a confidence level of $1-\alpha/3$. The same kind of procedure is valid for the simultaneous confidence intervals.

Remark 3 *We estimate the 210+53+53=316 parameters (after compression) with 832×17=14144 data (that is about 45 data by parameter) which appears to be completely reasonnable.*

Global model
First of all the load curves are normalized by the total consumption, centered and standardized. Remind that in the global model, we study the entire load curves with 17520 points; there is no decomposition.

Compression phase We proceed as explained in Section 4.2 and the test of significance suggests us to keep 55% of the variance which corresponds to keep the 15 best frequencies (i.e. 30 best coefficients) and to a reduction factor of more than 580.

Regression phase Once the compression has been done, we proceed to the regression phase on the conserved coefficients. Finally, the estimates of the centered standardized load curves are given by:

$$\widehat{\mathrm{LC}}_k(t) = \sum_{j\in\mathcal{J}} \hat{a}_k^j \phi_j(t).$$

where $\mathcal{J}$ represents the set of the coefficients kept after compression. Since the dimension of the curves is too large leading to the failure of Genz algorithm, the confidence bands are constructed using the asymptotic result given in Section 3.3.

Comments and comparison of the results
We plot the different results in Figures **??** (load curve and estimations) and 5 (load curves and confidence bands) of individual 11 on $[200, 600]$ (for the sake of clarity) for the decomposed and the global models.
We remark that the estimation given by the decomposed model seems to be more accurate than the other one and fits better the load curve shape and so do the simultaneous confidence bands. More precisely, if we compare for example the medians of the errors, the decomposed model provides better estimates than the global model with 55% (model suggested by hard thresholding and significance tests techniques). On top of that, the decomposed model is more efficient, easy to compute and last but not least less consuming in terms of cost of simulation.

Figure 4: Comparison between the two models on the interval [200,600] (load curve in thick line, estimations in dashed line for the decomposed model and in solid line for the global model)

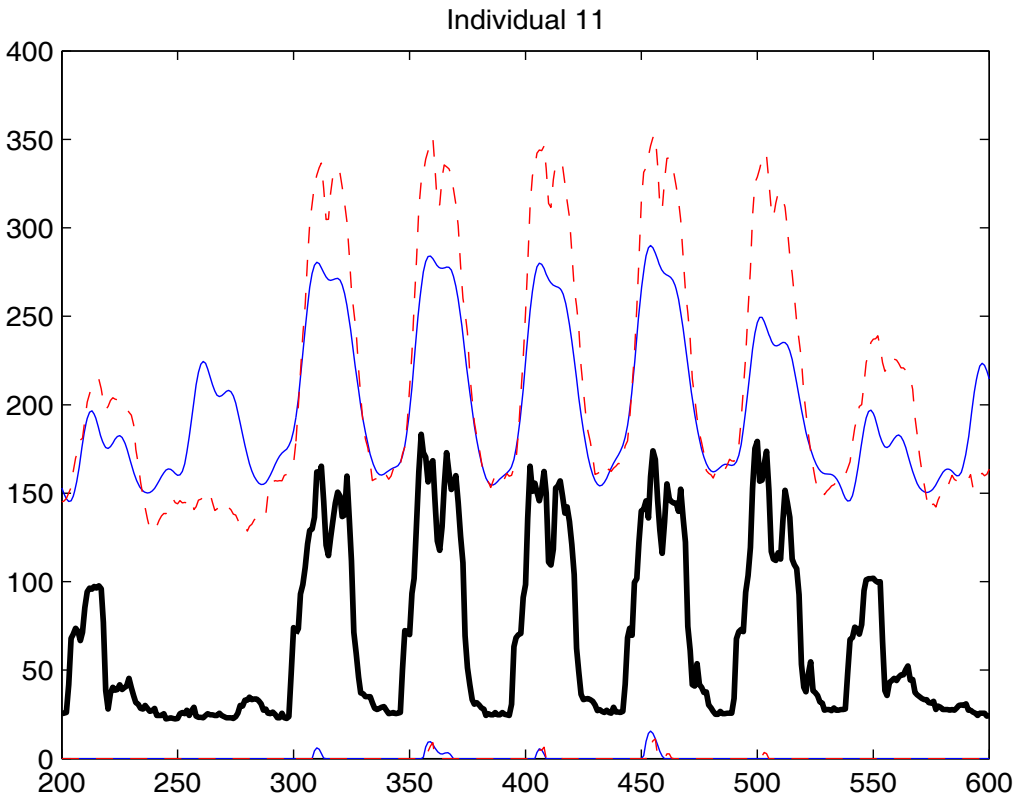

Figure 5: Comparison between the two models on the interval [200,600] (load curve in thick line, simultaneous confidence bands in dashed line for the decomposed model and in solid line for the global model)

Table 2: Errors for the test sample

Models	Median error	Mean error ± 5% conf. interval	Error range
Decomposed model	0.0826	0.1218 ± 0.0260	[0.0077,0.6860]
Global model	0.1047	0.1335 ± 0.0264	[0.0071,0.7185]

Table 3: Covering rates for the test sample

Model	SCR(1.00)	SCR(0.99)	SCR(0.95)	SCR(0.90)
Decomposed	56/85	82/85	84/85	85/85
Global	49/85	73/85	83/85	84/85

5.2 Test sample

We summarize the different results obtained for the test sample in Table 2 and Table 3. In terms of error, the decomposed model provides better results than the global model (corresponding to a gain about 0.27% for the median error). Concerning the covering ratios, remind that SCR(ρ) represents the ratio of individuals having 100ρ% of their curve in the SCB. For instance, 56 out of the 85 individuals of the test sample have their whole curve contained in the SCB for the decomposed model; 83 have 95% of their curve in the SCB for the global model.

We have to pay somewhere all the assumptions we have made to get our confidence bands. Table 3 shows that if we tolerate the curve to be out of the bound in 1% of the case, we get for the test sample a covering rate 82/85 which is consistent with the nominal value. In our opinion it justifies a posteriori our assumptions.

6 Conclusion and prospects

We have presented a method which is new in the sense that it is a combination of Sidak's inequality and numerical computation for Gaussian vectors. The method is easy to implement using few line Matlab programs. It has shown to be efficient in our problem of curve prediction for load curves. Moreover there is no doubt that the techniques developed here can be applied to many situations when one wants to predict a set of curves using a basis of functions.

References

[1] Azaïs, J.-M. and Genz, A. (2009). *Computation of the distribution of the maximum of stationary Gaussian sequences and processes*. Working paper.

[2] Azaïs, J.-M. and Mercadier, C. (2005). *Asymptotic Poisson character of extremes in non stationary Gaussian models*. Extremes,6, no4, 301-318.

[3] Bickel, P.J. and Rosenblatt, M. (1973). *On some global measures of the deviations of density function estimates*. The Annals of Statistics, Vol.1, No.6, 1071-1095.

[4] Claeskens, G. and Van Keilegom, I. (2003). *Bootstrap confidence interval for regression curve and their derivatives*. Ann. Statist. Vol.31, No.6.

[5] Csörgo, S. and Horvath, L. (1986). *Confidence bands from censored samples*. Can. J. Stat., Vol.14, No.2, 131-144.

[6] Donoho, D. L., Johnstone, I. M., Kerkyacharian, G. and Picard, D. (1995). *Wavelet Shrinkage: Asymptopia?*. J. Royal Stat. Soc. Ser. B, Vol.57, No.2, 301-369.

[7] Efron, B. (1997). *The length heuristic for simultaneous hypothesis tests*. Biometrika, Vol.84, No.1, 143-157.

[8] Genz, A. (1992). *Numerical computation of multivariate normal probabilities*. J. Comput. Graph. Statist. 1, 141-149.

[9] http://www.math.wsu.edu/faculty/genz/homepage

[10] Hall, P. (1992). *On bootstrap confidence intervals in nonparametric regression*. Ann. Statist. 20, 695-711.

[11] Hall, P. (1992). *Effect of bias estimation on coverage accuracy of bootstrap confidence intervals for a probability density*. Statistics 22, 215-232.

[12] Hall, P.G., Hyndman, R.J. and Fan, Y. (2004). *Nonparametric confidence intervals for receiver operating characteristic curves*. Biometrika 91, 743-750.

[13] Hotelling, H. (1939). *Tubes and spheres in n-spaces, and a class of statistical problems*. AM. J. Math. 61, No.2, 440-460.

[14] Kendall, W.S., Marin, J.M. and Robert, C.P. (2004). *Brownian confidence bands on Monte Carlo output*. Rapport de recherche INRIA.

[15] Lagnoux, A. (2008). *Estimation de courbes de charge et bandes de confiance simultanées*. Rapport de recherche EDF.

[16] http://www.lsp.ups-tlse.fr/Fp/Lagnoux/.

[17] Leadbetter, M. R. and Lindgren, Georg and Rootzén, Holger (1983). *Extremes and related properties of random sequences and processes*. Springer Series in Statistics, Springer-Verlag.

[18] Macskassy, S.A., Provost, F. and Rosset, S. (2005). *ROC confidence bands : An Empirical Evaluation*. Proceedings of the 22nd and International Conference on Machine Learning (ICML). Bonn, Germany.

[19] Miller, R. G., Jr. (1981). *Simultaneous Statistical Inference*. New York. Springer Verlag.

[20] Misiti, M., Misiti, Y., Oppenheim, G., Poggi, J.M. (1996). *PREVISOR : un logiciel de prévision non paramétrique. Notice d'utilisation succincte*. EDF Report.

[21] Misiti, M., Misiti, Y., Oppenheim, G., Poggi, J.M. (2007). *Analyse de l'apport des ondelettes pour la prévision des séries chronologiques dans le contexte électrique*. EDF Report.

[22] Naiman, D. Q. (1986). *Conservative bands in curvilinear regression*. Ann. Statist. 14, No.3, 896-906.

[23] Newton, J. and Pagano, M. (1984). *Simultaneous confidence bands for autoregressive spectra*. Biometrika 71, 197-202.

[24] Scheffé, H. (1959) *The analysis of variance*. John Wiley & Sons.

[25] Sidak, Z. (1967). *Rectangular confidence regions for the means of multivariate normal distributions*. J. Amer. Stat. Assoc. 62,626-633.

[26] Tukey, J. (1953). *The problem of multiple comparisons*. Unpublished manuscript.

[27] Zouhourian-Saghiri, L., Kobilinsky, A., Gillon, Y. and Gagnepain, C. (1983). *Loi d'usure mandibulaire chez Locusta Migratoria (Orthopt. Acrididae). Son utilisation pour la datation des ailés*. Ann. Soc. ent. Fr.(N.S.), 19(3), 335-352.