Multilevel functional principal component analysis of façade sound insulation data

Raffaele Argiento[‡], Pier Giovanni Bissiri[‡,*], Antonio Pievatolo[‡] and Chiara Scrosati[§,†]

[‡]IMATI–CNR, via Bassini 15, 20133 Milan, Italy

[§]ITC–CNR, S. Giuliano Milanese (MI), Italy

Abstract

This work analyses the data from an experimental study on façade sound insulation, consisting of independent repeated measurements executed by different laboratories on the same residential building. Mathematically, data can be seen as functions describing an acoustic parameter varying with the frequency. The aim of this study is twofold. On one hand, considering the laboratory as the grouping variable, it is important to assess the within and between group variability in the measurements. On the other hand, in building acoustics it is known that sound insulation is more variable at low frequencies (from 50 to 100 Hz), compared to higher frequencies (up to 5000 Hz), and therefore a multilevel functional model is employed to decompose the functional variance both at the measurament and at the group level. This decomposition also allows for the ranking of the laboratories on the basis of their measurement variability and their different performances at both the low frequencies (relative high variability) and over the whole spectrum. The former ranking is obtained via the principal component scores and the latter via the functional depth.

Keywords: Facade sound insulation; Repeatability and reproducibility; Multilevel functional data analysis; Functional depth; Bayesian functional regression.

[*]Corresponding author.

[†]This study is based on a research conducted in a full-scale experimental building at the Construction Technologies Institute of the National Research Council of Italy (ITC-CNR) and sponsored by the Lombardy Region (Scrosati et al. 2013, Scamoni et al. 2009).

1 Introduction

Building acoustic tests on samples presumably made of the same materials, in identical conditions, generally do not give the same results. This condition is due to inevitable errors (systematic and random) in test procedures, caused by the difficulties in controlling the several factors that influence the test, such as acoustic instrumentation, acoustic method (microphones and sources position), context (regular rooms or semi-open space, with different dimensions), constructive details of the building (that could have effect on acoustic measurements) and workmanship and, concerning sound levels, influence of instrumentation operating conditions (repeat configuration).

In general, uncertainties should preferably be determined following the principles laid down in ISO/IEC Guide 98-3 (*Uncertainty of measurement – Part 3: Guide to the expression of uncertainty in measurement (GUM:1995). International Standard ISO/IEC Guide 98-3:* 2008). This Guide specifies a detailed procedure for the uncertainty evaluation that is based on a complete mathematical model of the measurement procedure. According to current knowledge, it seems impossible to formulate these models for the different quantities in building acoustics. Therefore, to determine the uncertainty of building acoustics measurements, concepts as repeatability (within variance) and reproducibility (between variance) are needed. The best way to study the repeatability and reproducibility of building acoustics field measurements is to carry out a Round Robin Test (RRT), which consists of independent measurements executed several times by different laboratories.

Each single measurement is summarized by an acoustic index, which can be considered as the response of a one-way random effects model. The usual ANOVA methodology provides an estimate of a reference value for the acoustic index itself, by the general mean of all measurements, along with estimates of the repeatability and reproducibility errors, by the appropriate roots of mean squares. This univariate analysis is of interest in itself because the uncertainty of field measurements, particularly in facade sound insulation, has not been investigated comprehensively in the past (as pointed out by Scrosati et al. 2015). In this work, we examine a more general approach which aims at extracting all the information contained in the measurements, instead of analysing single-index summaries.

Each RRT measurement yields a value for each one of 21 given frequency bands, and therefore it provides a discrete sampling from a function defined on a frequencies range. In other words, the data are functions, and therefore the most suitable statistical framework is what is called *functional data analysis* (FDA). Such term, which has been introduced by Ramsay and Silverman, defines a branch of statistics which has received an increasing interest in the last years. It regards data that are records of values reflecting a smooth variation so that their sampling intervals could be as small as desired. It is therefore natural to assume the existence of smooth functions (*functional data*) giving rise to the observed data. This is the main difference between functional data analysis and multivariate data analysis, although the two branches of statistics are closely related because functional data analysis deals with highly multivariate objects. In other words, FDA differs from multivariate data analysis due to the intrinsic order of the observations, which can be related to time, space, as for time series or images, or frequency, which is our case. For an overview of the topic, see the two monographs Ramsay & Silverman (2002, 2005) and citations provided therein.

Two general areas can be considered in the framework of FDA: functional linear models (which include functional analysis of variance, functional ANOVA) and functional principal component analysis (FPCA). Functional ANOVA is useful to obtain functional decomposition when a natural hierarchy of units is present. A rich collection of inferential methods for functional linear models can be found in the statistical literature, as a result of a relevant methodological research effort. For instance, Brumback & Rice (1998) introduce smoothing spline models for nested and crossed curves, Guo (2002) discusses functional mixed effects models, Morris et al. (2003), Morris & Carroll (2006) propose Baysian wavelets models, Bigelow & Dunson (2007) introduce Bayesian adaptive regression splines, Baladandayuthapani et al. (2008) analyze Bayesian models for spatially correlated functional data.

The second general area, which is FPCA, is aimed to identify the most important modes of variation in the data and it can also be useful for dimension reduction. In a nutshell, a suitable orthogonal basis is built so that the coordinates of each data represent its amount of variation in the functional principal directions. For further details about FPCA, see Ramsay & Dalzell (1991), Silverman (1996), James et al. (2000), Yao et al. (2005), and also Hall & Hosseini-Nasab (2006), besides Ramsay & Silverman (2005).

Both these two general areas provide appropriate tools to fit the acoustic data in this paper. Indeed, we expect higher variability on some sub–spectra of the frequency: as known in the engineering literaturate, acoustic indexes are more variable at the lowest frequencies; higher variability is also expected at some other specific frequency bands, due to the framework of the experiment (i.e. some particular features of the experimental building). FPCA is suitable to capture this behaviour. On the other hand, our data are characterized by a natural hierarchy intrinsically implied by the RRT, which makes the functional ANOVA the routine choice. In fact, this tool decomposes the total covariance estimating the within laboratory and the between laboratory variability. In this way, one can analyze variability in terms of reproducibility and repeatability as generally done in the engineering literaturate, but in a functional framework. In ligth of these considerations, a model is needed that is capable of tackling both aspects of the data. Such model should extend FDA within multilevel functional analysis. This need has motivated a number of paper. See, for instance, Baladandayuthapani et al. (2008), Di et al. (2009), Guo (2002), Morris et al. (2003), Morris & Carroll (2006), Crainiceanu et al. (2009), Staicu et al. (2010).

Here, following Di et al. (2009), a Bayesian *multilevel functional principal component analysis* (MFPCA) is implemented (see also Crainiceanu & Goldsmith 2010). Indeed, as argued by Crainiceanu & Goldsmith (2010), Bayesian analysis is particularly appropriate when dealing with mixed models representations. Moreover, Bayesian simulation algorithms work very well in the context of FDA due to the orthogonality of the principal component bases. Another strenght of Di et al. (2009)'s approach is that the Bayesian estimation can be obtained through a Gibbs sampler carried out within JAGS (Plummer et al. 2003), so that the implementation effort is light.

Finally, in this work, we would like to use the estimates under MFPCA to rank the laboratories according to their performance. Such performances are represented by the laboratory–specific mean functions. In particular, to order curves from the center outwards the depth concept is used. We extend the modified band depth of Sun & Genton (2011) to a probabilistic setting, so that the posteriors of the depths of the laboratory–specific means are correctly defined and computed. Moreover, depths will be also used to define the credible posterior bounds for the functional parameters of interest.

The outline of this paper is the following: Section 2 describes our data, Section 3 and 4 introduce FPCA and MFPCA, respectively, Section 5 is about functional depths, Section 6 reports our data analysis, Section 7 presents the robustness analysis, Section 8 compares the point and functional approaches, and Section 9 draws conclusions of the analysis.

2 Acoustic data

The building under test is an existing experimental building located at ITC-CNR headquarters, made of prefabricated concrete panels. The building element tested is a prefabricated concrete façade with a 4 mm single glazing wood-aluminium frame window with a MDF (Medium Density Fibreboard) shutter box. The façade is situated at the first floor and its surface is 8.6 m^2. The receiving room is a rectangular room of 54.5 m^3 volume.

Each team measured the standardised level difference of façade $D_{2m,nT}$, which is the level difference in decibels, corresponding to a reference value of the reverberation time in the receiving room. The level difference is the difference, in decibels, between the outdoor sound pressure level 2m in front of the façade and the space and time averaged pressure level in the receiving room.

Nine teams, coordinated by ITC-CNR were involved in the RRT, each of them operating with its own procedures and equipment. In a first approach of this study (Scrosati et al. 2015), a laboratory that showed a significant presence of stragglers and outliers was excluded. Even if it turned out that there was nothing wrong with the microphones and the measurement instrumentation, it was found that the differences between including and excluding this laboratory were remarkable. As the method was correctly followed, the presence of stragglers and outliers, without any other physical explanation, can only be attributed to an external event. Referring to this team, it turned out that the most plausible explanation of an external event was a not perfectly closed door.

Therefore the data consist of forty measurements of $D_{2m,nT}$: five measurements performed by each one of the eight teams involved in the RRT. Each measurement is made of 21 observed values at the 21 one-third/octave bands frequencies in Hz: 50; 63; 80; 100; 125; 160; 200; 250; 315; 400; 500; 630; 800; 1000; 1250; 1600; 2000; 2500; 3150; 4000; 5000. Each measurement

can be considered as the discretization of a function on the interval [50, 5000]. See Fig. 1. In other words, these measureaments are regarded as multilevel functional data, which is one motivation for the multilevel principal component analysis carried out in this paper. We remark again that using this methodology, it will be possible to evaluate the reproducibility and repeatability errors (as between and within laboratory errors, respectively) and to assess the effect of individual laboratories over the considered frequency range.

3 Introduction to functional principal component analysis

To be self–contained, we now recall the mathematical theory on which the statistical models used in this paper are based. The mathematical basis for functional principal component analysis consists of two important results in functional analysis: Mercer's representation theorem and Karhunen–Loéve expansion. Let I a closed interval in $\mathbb{R}$. Recall that $L^2(I)$ is the space of Lebesgue square–integrable functions on I, i.e. of all functions f on I such that $\int f(x)^2 \mathrm{d}x$ exists and is finite. Moreover, $L^2(I)$ is a Hilbert space with inner product $\langle f, g\rangle = \int f(t)g(t)\mathrm{d}t$, and norm $\|f\| = \int_I |f(t)|^2 \,\mathrm{d}t$. Let $X = \{X(t) : t \in I\}$ be a random element of $L^2(I)$ such that

$$\mathbb{E}\{\|X\|^2\} = \mathbb{E}\left\{\int_I X^2(t)\mathrm{d}t\right\} < \infty.$$

Moreover let $\mu(t) = \mathbb{E}\{X(t)\}$ and $K(s,t) = \mathrm{Cov}\{X(s), X(t)\}$, $t, s \in I$, be the *mean function* and the *covariance function*, respectively. The *covariance operator* C is defined as:

$$\begin{aligned} C(f)(t) &= \mathbb{E}\{\langle X - \mu, f\rangle (X(t) - \mu(t))\} \\ &= \int_I K(s,t) f(s)\mathrm{d}s, \end{aligned}$$

for every function f in $L^2(I)$ and $t \in I$. By Mercer's theorem, the following spectral decomposition of the covariance function holds true:

$$K(s,t) = \textstyle\sum_{k=1}^{\infty} \lambda_k \varphi_k(s)\varphi_k(t), \tag{1}$$

where φ_k are the eigenfunctions of C and $\lambda_1 \geq \lambda_2 \geq \ldots$ the corresponding eigenvalues, i.e. the ϕ_k's are orthonormal functions in L^2 such that $C(\varphi_k) = \lambda_k \varphi_k$, for $k \geq 1$. The functions

φ_k's are the *principal components* of X. Indeed, (1) is equivalent to:

$$C(f)(t) = \int \left(\sum_{k=1}^{\infty} \lambda_k \varphi_k(s)\varphi_k(t)\right) f(s)\mathrm{d}s = \sum_{k=1}^{\infty} \lambda_k \left(\int f(s)\varphi_k(s)\mathrm{d}s\right) \varphi_k(t).$$

The following is the *Karhunen–Loéve representation* for the random element X:

$$X(t) = \mu(t) + \sum_{k=1}^{\infty} \xi_k \varphi_k(t), \tag{2}$$

where

$$\xi_k = \int \{X(t) - \mu(t)\}\varphi_k(t)\mathrm{d}t \tag{3}$$

are zero–mean uncorrelated random variables such that $\mathrm{Var}(\xi_k) = \lambda_k$, $k \geq 1$, which are called *principal components scores*. The Karhunen–Loéve representation (2) entails that:

$$\mathrm{Var}(\xi_k) = \lambda_k, k \geq 1,$$
$$\int \mathrm{Var}\{X(t)\}\mathrm{d}t = \sum_{k\geq 1} \mathrm{Var}(\xi_k) = \sum_{k\geq 1} \lambda_k.$$

4 Multilevel functional principal component analysis

The method we apply here is multilevel functional principal component analysis, which has been introduced by Di et al. (2009). See also Crainiceanu & Goldsmith (2010). Therefore we refer to the functional random effects model:

$$X_{i,j}(t) = \mu(t) + Z_i(t) + W_{i,j}(t),$$

where $X_{i,j}(t)$ is the j-th measurement of the i-th laboratory related to the t Hz frequency, for $i \in I = \{1, \ldots, 8\}$, $j \in J = \{1, \ldots, 5\}$, t ranges in the interval $[50, 5000]$, $\mu(t)$ is the overall mean function, $Z_i(t)$ is the laboratory–specific deviation from the mean, and $W_{i,j}(t)$ is the residual laboratory– and measurement–specific deviation from the laboratory–specific mean. The random processes $(Z_i(t))_t$, $(W_{l,j}(t))_t$ are mean–zero pairwise uncorrelated processes, for $i, l \in I$, $j = 1 \in J$. Let $K_T(t,s) = \mathrm{Cov}(X_{i,j}(t), X_{i,j}(s))$ be the total covariance, $K_B(t,s) = \mathrm{Cov}(Z_i(t), Z_i(s))$ the between covariance and $K_W(t,s) = \mathrm{Cov}(W_{i,j}(t), W_{i,j}(s))$ the within covariance. Of course, $K_T(t,s) = K_B(t,s) + K_W(t,s)$.

Level one $\phi_k^{(1)}(t)$ and level two $\phi_l^{(2)}(t)$ principal components are the eigenfunctions that arise in the spectral decomposition of the between and within covariance function, respectively:

$$K_B(s,t) = \sum_{k=1}^{\infty} \lambda_k^{(1)} \varphi_k^{(1)}(s)\varphi_k^{(1)}(t),$$
$$K_W(s,t) = \sum_{l=1}^{\infty} \lambda_l^{(2)} \varphi_l^{(2)}(s)\varphi_l^{(2)}(t),$$

where $\lambda_k^{(1)}$ and $\lambda_l^{(2)}$, $k, l \geq 1$ are level 1 and level 2 eigenvalues, respectively. Note that $\{\varphi_k^{(1)} : k \geq 1\}, \{\varphi_l^{(2)} : l \geq 1\}$ are orthonormal bases in $L^2(T)$, but are not required to be mutually orthogonal.

The Karhunen-Loéve expansions of the functions Z_i and $W_{i,j}$ are:

$$Z_i(t) = \sum_{k=1}^{\infty} \xi_{i,k}\varphi_k^{(1)}(t), \qquad W_{i,j}(t) = \sum_{l=1}^{\infty} \zeta_{i,j,l}\varphi_l^{(2)}(t), \tag{4}$$

for $i \in I, j \in J$. The corresponding eigenvalues are denoted by $\lambda_k^{(1)}$ and $\lambda_l^{(2)}$, $k, l \geq 1$, respectively. Level one and level two principal component scores are $\xi_{i,k} = \int \phi_k^{(1)}(t) Z_i(t) \mathrm{d}t$ and $\zeta_{i,j,l} = \int \phi_l^{(2)}(t) W_{i,j}(t) \mathrm{d}t$, respectively. Recall that $\mathrm{Var}(\xi_{i,k}) = \lambda_k^{(1)}$ and $\mathrm{Var}(\zeta_{i,j,l}) = \lambda_l^{(2)}$, for every i, j, k, l. Moreover,

$$\int \mathrm{Var}(X_{i,j}(t))\mathrm{d}t = \sum_{k\geq 1} \lambda_k^{(1)} + \sum_{l\geq 1} \lambda_l^{(2)},$$

for every i, j. For each k and l, the proportions of the variance explained by level one k–th and level two l–th principal component are $\lambda_k^{(1)} / \sum_{k\geq 1} \lambda_k^{(1)}$ and $\lambda_l^{(2)} / \sum_{l\geq 1} \lambda_l^{(2)}$, respectively. A natural measure of variance explained by between laboratories variability is the ratio:

$$\frac{\int \mathrm{Var}(Z_i(t))\mathrm{d}t}{\int \mathrm{Var}(Z_i(t))\mathrm{d}t + \int \mathrm{Var}(W_{i,j}(t))\mathrm{d}t} = \frac{\sum_{k\geq 1} \lambda_k^{(1)}}{\sum_{k\geq 1} \lambda_k^{(1)} + \sum_{l\geq 1} \lambda_l^{(2)}}. \tag{5}$$

This is called the *functional intra–cluster correlation*. Indeed, it measures how measurements made in the same laboratory are close to each other. Such value can be easily estimated plugging in the estimated values of the eigenvalues.

Following Di et al. (2009), we estimate the covariance functions assuming that we observe the noisy data:

$$Y_{i,j}(t) = X_{i,j}(t) + \varepsilon_{i,j}(t)$$

where $\varepsilon_{i,j}(t)$ is a white noise process with variance σ^2.

Let

$$G_T(t,s) = \mathrm{Cov}(Y_{i,j}(t), Y_{i,j}(s)),$$
$$G_B(t,s) = \mathrm{Cov}(Y_{i,j}(t), Y_{i,k}(s)).$$

Hence

$$G_T(t,s) = K_T(t,s) + \sigma^2 \mathbb{I}_{\{t=s\}}(t,s),$$
$$G_B(t,s) = K_B(t,s).$$

At this stage, we detail our procedure without underlying the minor differences with Di et al. (2009)'s approach:

1. By penalized spline smoothing, we obtain the estimate $\hat{\mu}(t)$ for $\mu(t)$;
2. We obtain the estimates $\hat{G}_T(t_s, t_r)$, $\hat{G}_B(t_s, t_r)$ by the methods of moments, for $s, r = 1, \ldots, 21$, being $t_1, \ldots, t_{21}$ the one-third/octave bands frequencies;
3. We obtain the estimate $\hat{K}_B(t,s)$ by smoothing $\hat{G}_B(t_s, t_r)$ (we fit semiparametric regression models using the mixed model representation of penalized splines, relying on the package *SemiPar* of *R* software);
4. We obtain the estimate $\hat{K}_T(t,s)$ by smoothing $\hat{G}_T(t_s, t_r)$ (as in the previous step) for $t_s \neq t_r$, i.e. dropping diagonal elements;
5. We obtain the estimate $\hat{K}_W(t,s) = \hat{K}_T(t,s) - \hat{K}_B(t,s)$ by difference;
6. We use eigenanalysis on $\hat{K}_B(t,s)$ to obtain $\hat{\varphi}_k^{(1)}(t)$, and $\hat{\lambda}_k^{(1)}$;
7. We use eigenanalysis on $\hat{K}_W(t,s)$ to obtain $\hat{\varphi}_l^{(2)}(t)$ and $\hat{\lambda}_l^{(2)}$ (remark: since $\hat{K}_W$ is obtained by difference, it cannot be positive definite; we handle this problem by trimming eigenvalue–eigenvector pairs with negative eigenvalue – see, for instance, Hall et al. 2008).

Then, we assess the number of level one principal components to keep as following:

$$K_1 = \min\{k : \rho_k \geq 0.99, \hat{\lambda}_k^{(1)} < 1/21\}, \tag{6}$$

where

$$\rho_k = \sum_{h=1}^{k} \hat{\lambda}_h^{(1)} / \sum_{h \geq 1} \hat{\lambda}_h^{(1)}.$$

In the same way, we employ $\hat{\lambda}_l^{(2)}$ to assess the number K_2 of level two principal components to keep.

Recall that $\{\varphi_k^{(1)} : k \geq 1\}, \{\varphi_l^{(2)} : l \geq 1\}$ are not required to be mutually orthogonal. This makes it more difficult to estimate the scores for the multilevel model than for the basic model introduced in Section 3. Indeed, for that model, it is straightforward to estimate scores on the basis of (3) simply by approximating the integral:

$$\int \{X_i(t) - \hat{\mu}(t)\} \hat{\varphi}_k(t) \mathrm{d}t$$

where $\hat{\varphi}_k(t)$, $k = 1, \ldots, K$, are the estimated principal components. Instead, in the multilevel model, as shown in (4), eigenvalues are functions of the processes $Z_i(t)$ and $W_{i,j}(t)$, which are not directly observed.

To estimate eigenvalues and scores we employ the following Bayesian model:

$$Y_{i,j}(t) = \mu(t) + \sum_{k=1}^{K_1} \xi_{i,k} \varphi_k^{(1)}(t) + \sum_{l=1}^{K_2} \zeta_{i,j,l} \varphi_l^{(2)}(t) + \varepsilon_{i,j}(t),$$

where we plug in the estimates for $\mu(t)$ and $\varphi_k^{(1)}(t)$ and $\varphi_l^{(2)}(t)$. The priors for the scores make them independent centered normal random variables, with variances $\tilde{\lambda}_k^{(1)}$ and $\tilde{\lambda}_l^{(2)}$ for the k–th level one and the l–th level two scores, respectively. The hyperprior for the hyperparameters $\tilde{\lambda}_k^{(1)}$ and $\tilde{\lambda}_l^{(2)}$ make them independent inverse gamma random variables with means equal to the obtained eigenvalues of the between and within covariance functions, i.e. $\hat{\lambda}_k^{(1)}$ and $\hat{\lambda}_l^{(2)}$, $1 \leq k \leq K_1, 1 \leq l \leq K_2$, and the same variance σ_λ^2. The random variables $\varepsilon_{i,j}(t)$ are i.i.d. $N(0, \sigma^2)$, where the prior for the nugget effect σ^2 is an inverse gamma distribution with a large variance (to make it substantially non–informative). We denote by $\lambda_k^{*(1)}$, $k = 1, \ldots, K_1$, and by $\lambda_l^{*(2)}$, $l = 1, \ldots, K_2$, the Bayesian estimates for level one and two eigenvalues, respectively.

5 Functional depths

To rank functional data, several measures of centrality, called functional depths, have been introduced (Fraiman & Muniz 2001, Lopez-Pintado & Romo 2009, Sun & Genton 2011, see

for instance). Among them, we shall refer to the *modified band depth*, which has been introduced by Lopez-Pintado & Romo (2009) and also studied by Sun & Genton (2011). The basic idea is the band depth concept, which relies on a graph–based approach. The band $B(y_1, y_2)$ in $I \times \mathbb{R}$ delimited by two given curves $y_1(t)$ and $y_2(t)$ is $\{(t, y(t)) : \min\{y_1(t), y_2(t)\} \leq y(t) \leq \max\{y_1(t), y_2(t)\}\}$. Given a random curve, i.e. a stochastic process, $Y(t)$, the *band depth* for a given curve $y(t)$, is $P(\{(t, y(t)) : t \subset \mathbb{R}\} \subset B(Y_1, Y_2))$, where $Y_1(t)$ and $Y_2(t)$ are two independent copies of the stochastic process $Y(t)$ generating the observations $y_1(t), \ldots, y_N(t)$. Here, we consider a more flexible definition based on the proportion of the domain I where a curve $y(t)$ is in the band. Formally, we consider the set:

$$A(y; y_1, y_2) = \{(t \in I : \min\{y_1(t), y_2(t)\} \leq y(t) \leq \max\{y_1(t), y_2(t)\}\}.$$

Hence, the *modified band depth* for a given curve $y(t)$ can now be defined as:

$$\text{MBD}(y) = \mathbb{E}\left(\frac{\lambda(A(y; Y_1, Y_2))}{\lambda(I)}\right), \tag{7}$$

where λ is the Lebesgue measure. Letting T be an uniform random variable on I independent of $Y(t)$, (7) becomes:

$$\text{MBD}(y) = P(\min\{Y_1(T), Y_2(T)\} \leq y(T) \leq \max\{Y_1(T), Y_2(T)\}).$$

The depth defined by (7) is the probabilistic generalization of the sample modified band depth introduced by Lopez-Pintado & Romo (2009), which is:

$$\text{MBD}_N(y) = \frac{2}{N(N-1)} \sum_{1 \leq i_1 < i_2 \leq N} \frac{\lambda(A(y; y_{i_1}, y_{i_2}))}{\lambda(I)}.$$

Through functional depths, Sun & Genton (2011) generalize to the functional data setting the concepts of quantiles, interquantile range and they introduce the functional boxplot. To this aim, they resort to the concept of central region introduced by Liu et al. (1999). If we transpose the idea of Liu et al. (1999) about Euclidean spaces to the functional setting, the *p–th central region*, for $0 < p < 1$, can be defined as:

$$C_p = \cap_t \{R(t) : P(R(t)) \geq p\},$$

where $R(t) = \{y \in L^2(I) : \text{MBD}(y) > t\}$ is the region enclosed by the contour of depth t. In other words, C_p is the smallest region enclosed by depth contours to a mass probability p.

Sun & Genton (2011) use the idea of Liu et al. (1999) to define the sample 50% central region and to built functional boxplots. Given the observed curves, $y_1, \ldots, y_N$, we can order them on the basis of their depths: $y_{[1]}, \ldots, y_{[N]}$. So, $y_{[1]}$ is the deepest (most central) curve or simply the median curve, $y_{[N]}$ is the most outlying curve, and in general $y_{[i]}$ denotes the sample curve associated with the the i–th largest band depth value. Hence, one can define the sample p–th central region as the band delimited by the most central fraction p of the observed curves, that is:

$$C_{N,p} = \left\{ (t, y(t)) : \min_{r=1,\ldots,[Np]} y_{[r]}(t) \leq y(t) \leq \max_{r=1,\ldots,[Np]} y_{[r]}(t) \right\} \tag{8}$$

If $p = 1/2$, one obtains the 50 % central region, which is the band delimited by the most deepest half of the observed curves. Such region is the analog to the "inter–quantile range" (IQR) and gives a useful indication of the spread of the central 50% region of the curve (Sun & Genton 2011). Moreover, as one could expect, it will be crucial to construct functional boxplots.

To define the 50% central region, one could consider an alternative and more intuitive approach, simply calculating the quantiles pointwise. Sun & Genton (2011) point out that such choice should be avoided. Indeed, in this way, the information about the curves' shapes would be lost. The median curve obtained putting together the medians point by point would be much more regular than and different from any observed curve. Moreover, the central curves obtained by pointwise boxplot would be narrower than those given by functional boxplots and fewer curves would be contained in the central region. Through functional depths, it is possible to treat each curve as one observation, which certainly is the most suitable approach in a functional data framework.

In our analysis, depths have been calculated to rank the estimated between effect functions $\hat{Z}_i(t)$ and the estimated within effect functions $\hat{W}_{i,j}(t)$. Moreover, we shall build functional boxplots associated to the posterior distribution of $Z_i(t)$, for $i = 1, \ldots, 8$.

6 Data analysis

In this section, we describe the results obtained by fitting the model illustraded in Section 4 to the acoustic data of Section 2.

We have estimated principal components and scores by the procedure described in Section 4, assessing $\sigma^2_\lambda = 10^3$ and the hyperparameters of the inverse–gamma distribution of σ^2 as 2 and 10^{-3}. In this way, we have obtained our smoothed data by means of the Karhunen–Loéve expansion. Figure 1 displays raw and smooth data (on the log scale as all the other graphs in this paper). As expected, high variability is quite apparent at low frequencies.

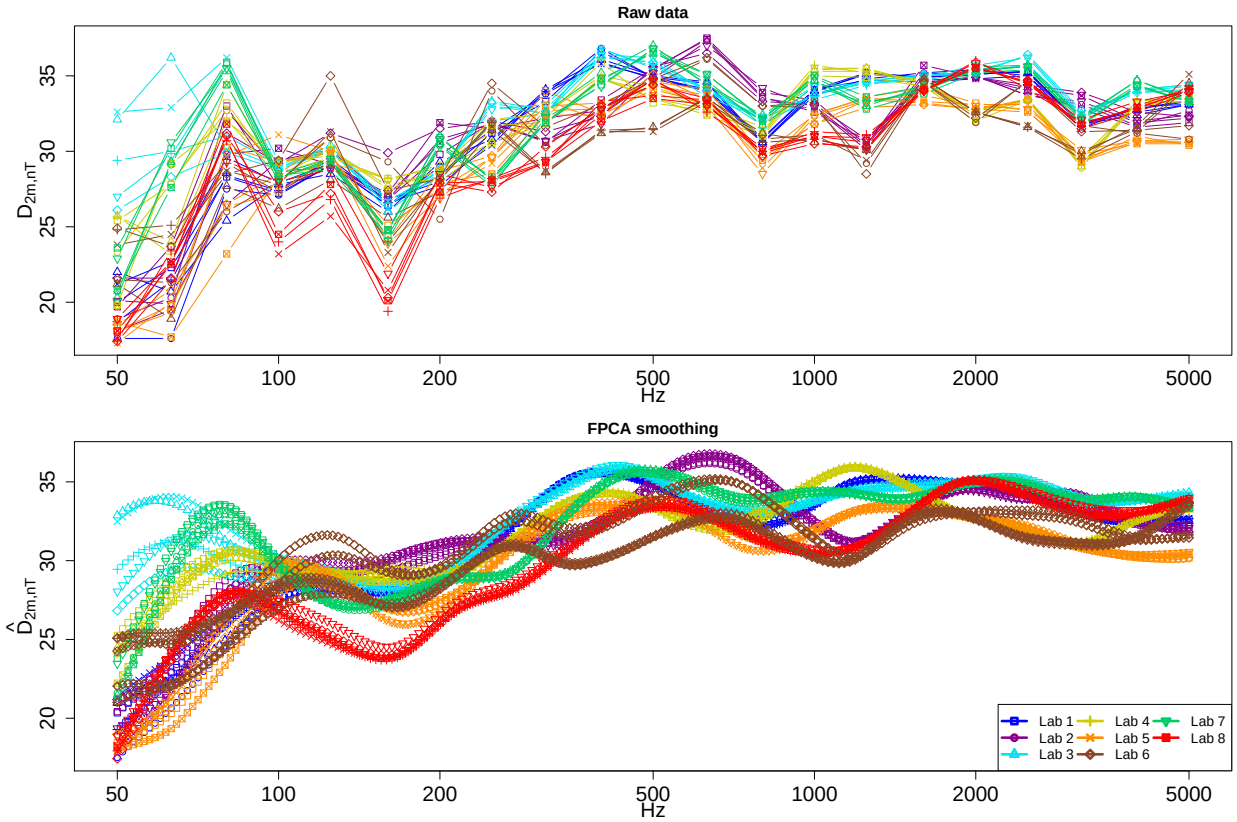

Figure 1: Raw and smoothed data

We have assessed the number of level one and two principal components to be retained, by stopping at the component where the percentage of variance explained is less than $1/T$, with $T = 21$, i.e. the number of grid points for each curve, and where the cumulative percentage of variance explained is greater than 99%, according to (6). A further constraint is that these numbers, for each level, must not be lower than four. We have obtained $K_1 = 6, K_2 = 5$. In other words, we have kept six components for level one and five for level two. The scores have been estimated using the Bayesian procedure described in Section 4.

We have estimated the intra–cluster correlation (5) by plugging in our Bayesian estimates

for the eigenvalues. We have obtained 0.885. In other words, 88.5 % of variability is attributable to the laboratory level variability. This is quite clear if we look at the Bayesian estimates of between, within and total variances (see Fig. 2), obtained as the diagonals of the corresponding estimates of the covariance operators

$$\tilde{G}_B(s,t) = \sum_{k=1}^{K_1} \hat{\lambda}_k^{(1)} \hat{\varphi}_k^{(1)}(s) \hat{\varphi}_k^{(1)}(t),$$
$$\tilde{G}_W(s,t) = \sum_{l=1}^{K_2} \hat{\lambda}_l^{(2)} \hat{\varphi}_l^{(2)}(s) \hat{\varphi}_l^{(2)}(t) + \hat{\sigma}^2 \mathbb{I}_{\{t=s\}},$$
$$\tilde{G}_T(s,t) = \tilde{G}_B(s,t) + \tilde{G}_W(s,t),$$

where $\hat{\sigma}^2$ is the estimated nugget effect. The between variance is always equal or greater

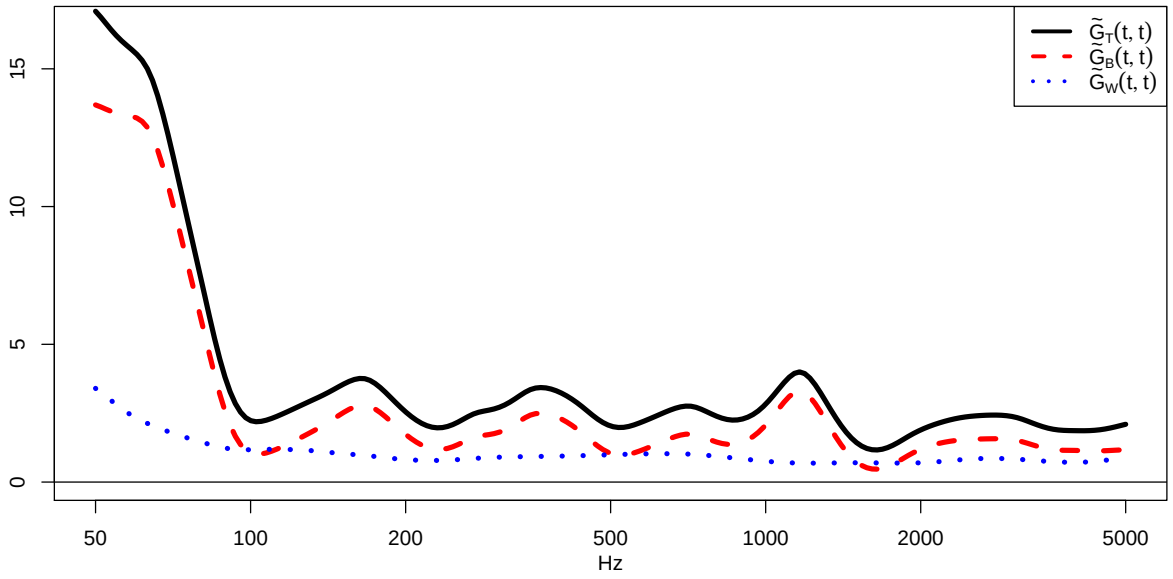

Figure 2: Between (red dashed line), within (blue dotted line) and total variance (black solid line).

than the within variance. Moreover, the between variance and the total variance have both a local maximum at 1163.6 Hz. In the first approach of this analysis (Scrosati et al. 2015), a high between-laboratories variation was observed at the 1250 Hz one-third octave band, corresponding to the critical frequency of the shutter box.

Figure 3 shows the estimates of the corresponding between and within correlation functions $\tilde{G}_T(t,t), \tilde{G}_B(t,t)$ and $\tilde{G}_W(t,t)$. We observe that the within correlation is generally low while the between correlation reaches higher values (> 0.8). Such peaks, represented by the darkest areas in the figure, are obtained around the diagonals and in the upper left and bottom right corners. This shows the tendency for those laboratories that overestimate (underestimate) at the low frequencies to behave similarly on the high frequencies, as it will be confirmed by the analysis of the first level-one principal component.

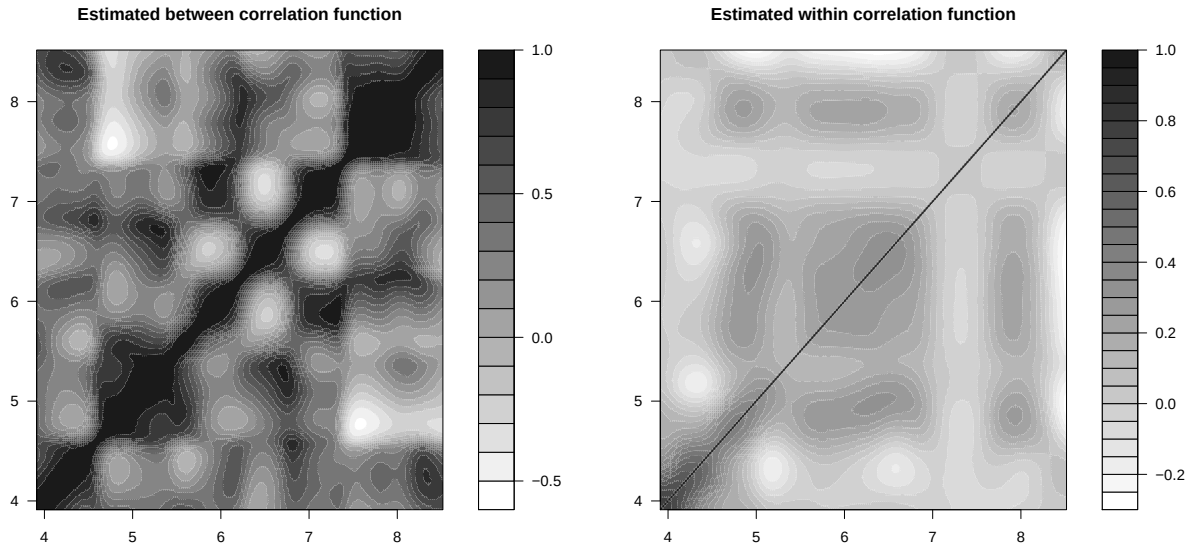

Figure 3: Estimated between and within correlation functions.

We shall now evaluate level-one principal components. Fig. 4 is a typical graph in functional data analysis: in the first row, one can find estimates of level-one principal components, whereas the graphs in the second row display the population mean function (solid line) and the functions obtained by adding and subtracting a suitable multiple of the eigenfunctions from the mean (the lines with the plus symbol and minus symbol markers, respectively). More precisely, we have plotted the function $\hat{\mu}(t)$ together with $\hat{\mu}(t) + \sqrt{\lambda_k^{*(1)}}\hat{\varphi}_k^{(1)}(t)$, $\hat{\mu}(t) - \sqrt{\lambda_k^{*(1)}}\hat{\varphi}_k^{(1)}(t)$, for $k = 1, \ldots, K_1$. More than half (53,15 %) of the between laboratories variability is explained by the first principal component. Such component is negative over the whole frequency range: it is referred to a negative shift of a laboratory–specific mean with respect to the general mean. It reaches a (negative) peak on the lowest frequencies whose

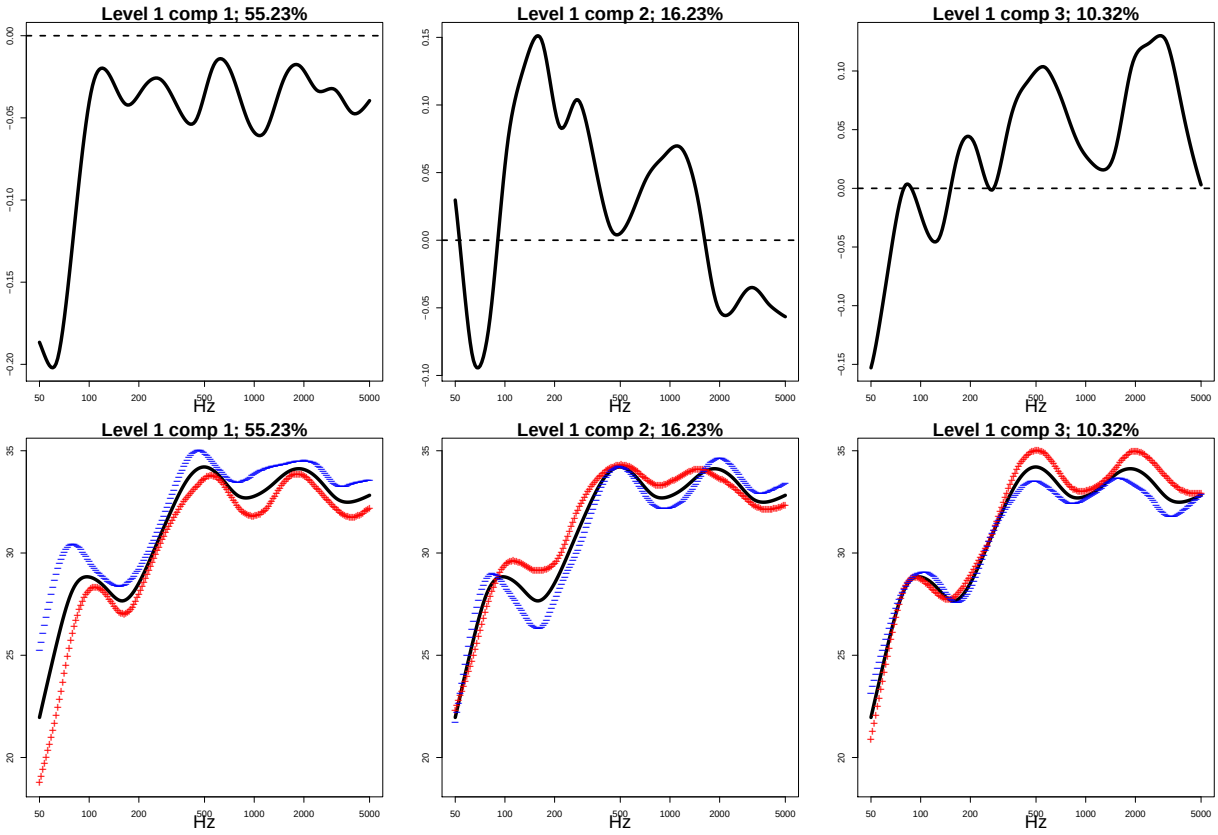

Figure 4: Level one principal components

magnitude is more than four times any other peak. This means that the greatest variability between laboratories will be found by heavily weighting the lowest frequencies, with only a light contribution from the other frequencies; the quantity $D_{2m,nT}$ is most variable across laboratories on the lowest frequencies, in short.

From the acoustic point of view, a possible explanation for the high values at the lowest frequencies (i.e. the 50, 63 and 80 Hz one-third octave bands) can be sought in the presence of the normal modes of vibration, which are orthogonal to the direction of wave propagation. For a room of the size of the receiving room, the analysis of these modes of vibration (see Scrosati et al. 2015) confirms that, at the first three one-third octave bands, the measured levels can be strongly influenced by the microphone position.

The second level–one principal component has a negative peak at 69.13 Hz and two positive one at 159.03 Hz and 1110.97 Hz, of which the first one is quite sharp. Therefore, measures made in laboratories for which the score $\xi_{i,2}$ is high will be much lower than average at 69.13

Hz and higher at 159.03 Hz and 1110.97 Hz. This type of variation explains 16.23 % of the between laboratories variation mainly in the central part of the frequency spectrum as seen in the central bottom panel of Figure 4.

The other level-one principal components have an oscillatory behaviour, which is difficult to interpret. This is a problem that often arises in functional principal components analysis, but it is not a concern for us, as the remaining components account for small proportions of the variation.

We can now look at level–two principal components in Fig. 5. Here, there is not so much

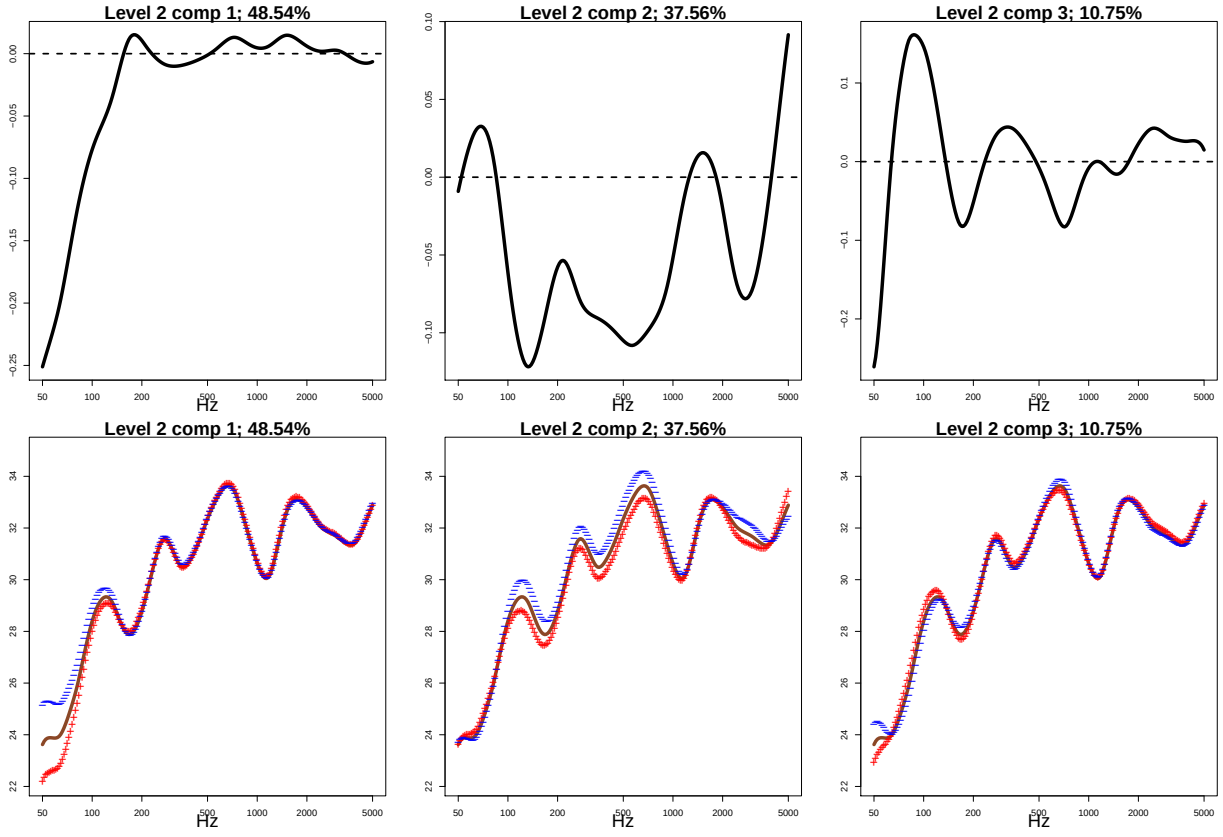

Figure 5: Level two principal components

difference about the portion of explained variability between the first and the second component. Indeed, the first and second one explain 48.54 % and 37.56% of the within laboratory variability, respectively. The first one has effect only on the lowest frequencies. In fact, it is close to zero for all frequencies greater than 155 Hz. Instead, the main effect of the second one regards frequencies ranging from 85.14 Hz to 1247.25 Hz and from 1848.46 Hz

to 3876.3 Hz, on which it accounts for a negative deviation of the single measure from the laboratory–specific mean. The remaining components again exhibit an oscillatory behaviour that is difficult to interpret, but, on the other hand, account for a small proportion of the within–laboratory variation, and therefore their interpretation is not a concern.

Let us now look at the posterior distributions of level-one principal component scores, which have been estimated by our Bayesian model. Fig. 6 displays the boxplots referred to their posterior quantiles. Let us recall that such scores express the weight given by first level one principal component to each laboratory, which is away from zero with negative values especially on the low frequencies. The analysis of the first principal component score allow us to rank laboratory performances when measuring low frequencies. We can see that laboratory three has a negative first principal component score, which is the highest in absolute value. Therefore, we expect that for this laboratory measured values are much higher than the general mean at the lowest frequencies. In other words, laboratory three seems to be responsible of the worst systematical error, with a tendency to overestimate $\mathrm{D}_{2m,nT}$. Other laboratories with negative scores are laboratory four and seven. Hence, these three laboratories generally overestimate the quantity on the low frequencies. On the contrary, laboratory five, six and eight generally underestimate the quantity on such frequencies. Laboratories with smallest scores in absolute values, such as laboratory one and two are able to provide more precise measureaments for the low frequencies than the other laboratories.

These conclusions are confirmed by the plot of the $\hat{Z}_i$ (Fig. 7): compare for instance laboratory one and three to other ones at the lowest frequencies.

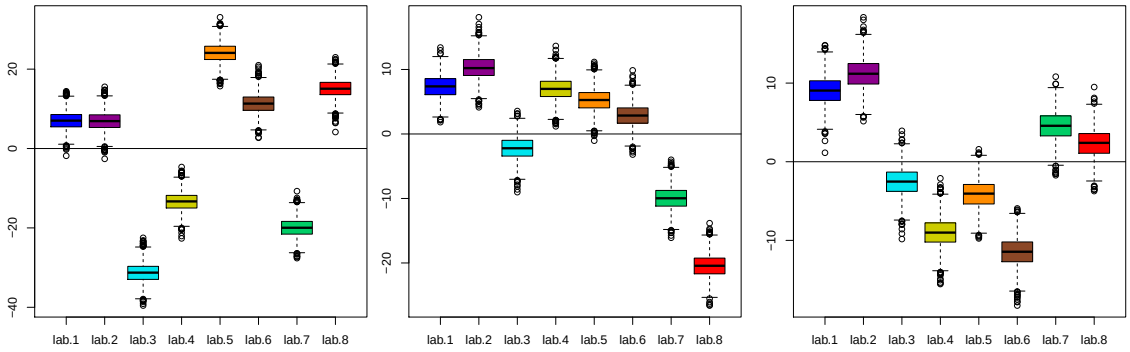

Figure 6: Estimated level one principal component scores $(\hat{\xi}_{i,1}, \hat{\xi}_{i,2}, \hat{\xi}_{i,3})$.

In Fig. 7, we have also obtained functional boxplots related to the posterior distributions of the deviation of the laboratory specific–means from the general mean. For each laboratory $i = 1, \ldots, 8$, our aim is to estimate the p–th central region C_p as defined in Section 5, where $Y(t)$ is replaced by $Z_i(t)$ with its posterior distribution. To this aim, we compute $C_{0.5,N}$ as in formula (8), where the observations $y_1, \ldots, y_N$ are the MCMC sample from the posterior of $Z_i(t)$. In this way, we obtain, for each laboratory $i = 1, \ldots, 8$, the area between the dashed lines, which is the 50% credible band of $Z_i(t)$. The border of the 50% central region is defined as the envelope representing the box in the classical boxplot. The dotted most external lines are the functional analog of the "whiskers" in the classical boxplot, obtained by inflating the envelope of the 50 % central region by 1.5 times its range. The points in Fig. 7 are the observed deviations of the laboratory–specific means from the general mean. Functional boxplots confirm what we have already observed, looking at level one scores, about the variability at the low frequencies. For instance, lab 3 and 5 deviate from the general mean noticeably, lab 3 upwards, lab 5 downwards, whereas lab 1 and 2 are the closest ones to the mean. Looking at Fig. 7, one can assess that performance of $Z_6(t)$ does not deviate much from zero, nevertheless $Z_6(t)$ does not fit quite well the point cloud of the observed deviations from the mean due to a high internal variability.

As one of the main goals of this paper, we now want to rank the global performances of the laboratories. To this aim, the best laboratory will be the one with the smallest laboratory–specific mean deviation $Z_i(t)$. Since these quantities are functions, we rank laboratories by the estimated depths of the Z_i's, as described in Section 5. The greater the depth the more central the curve. Fig. 8 summarizes the posterior distributions of depth for each laboratory. The specific mean deviation of laboratory one is the most central. This optimality comes out considering not only the posterior median of the depths but also the whole posterior distribution summarized through the boxplot. It is interesting to note that laboratory one is the laboratory who used the rotating microphone boom for the measurament of the space and time averaged pressure level in the receiving room, while all the other laboratories used the fixed microphones positions.

We can also use depths to compare measurements on the basis of their within laboratory variability. To this aim, we estimate depths of the estimates of $W_{i,j}$'s see Table 1. All

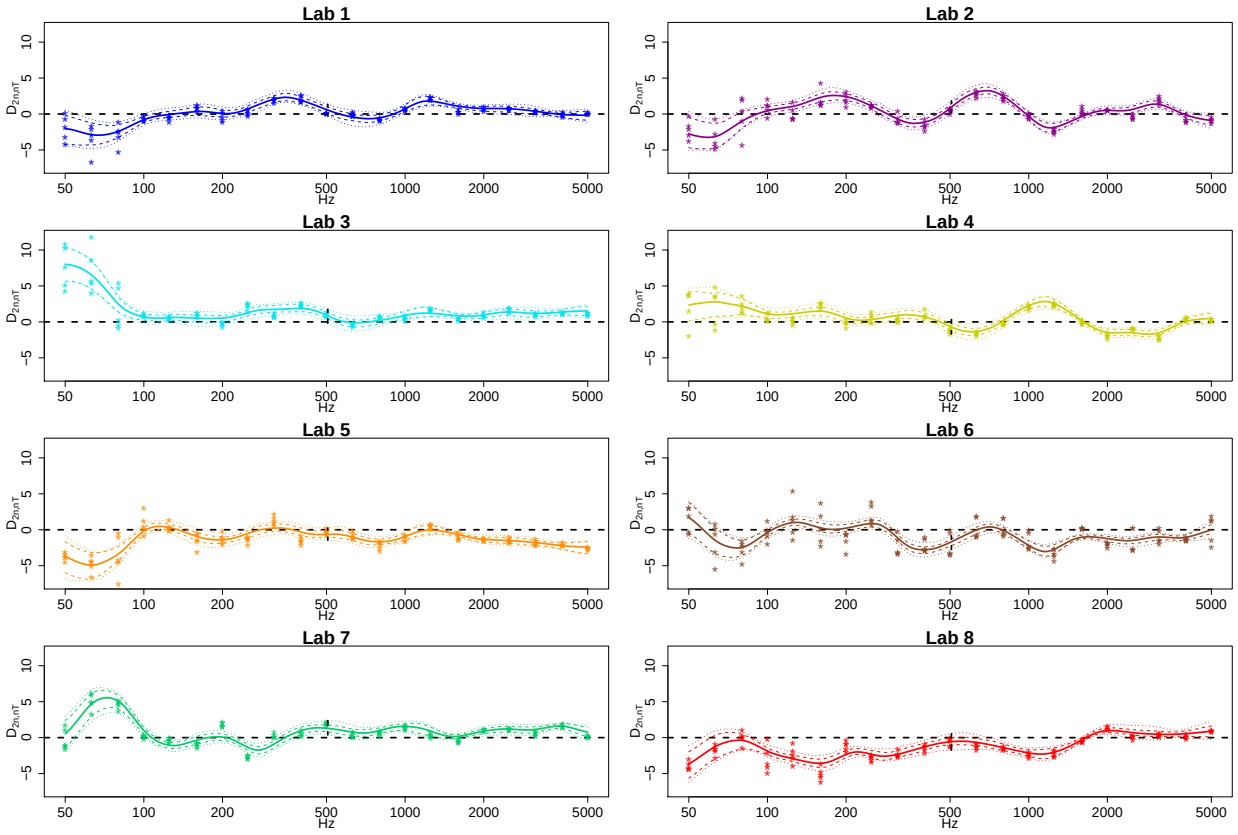

Figure 7: Functional laboratory deviation from the overall mean.

measuremetns of laboratory 6 have the lowest depth. Hence, laboratory 6 has the highest within variability. This confirms our comment about the fitting of $\hat{Z}_6(t)$ while discussing Fig. 7.

7 Robustness analysis

When we set up the Bayesian regression model introduced at the end of Section 4, one critical choice is the prior distribution for the mixed effect variances $\tilde{\lambda}_k^{(1)}$ and $\tilde{\lambda}_l^{(2)}$. The natural choice for the prior means are the eigenvalues of the covariance functions computed as reported at points 6 and 7 on page 9, while no prior information is available on the variance σ_λ^2. To deal with this lack of information, we have conducted a robustness analysis, comparing the different estimates for level–one and level–two eigenvalues obtained as σ_λ^2 varies in $\{10^j, j = 1, \ldots, 4\}$. Table 2 reports the eigenvalues of the between covariance

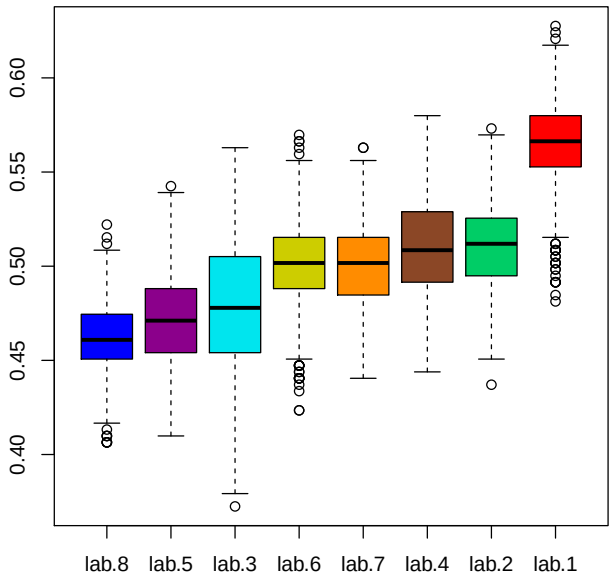

Figure 8: Depths of laboratory–specific means

Lab	6	6	6	6	6	3	2	5	8	2
	0.0989	0.1307	0.1441	0.1548	0.2078	0.2488	0.2656	0.2963	0.3149	0.3224
Lab	8	7	3	1	5	4	5	3	5	8
	0.3281	0.3395	0.3402	0.3436	0.3450	0.3499	0.3508	0.3526	0.3527	0.3579
Lab	2	7	2	1	3	8	3	4	1	5
	0.3639	0.3751	0.3880	0.3902	0.3997	0.4125	0.4174	0.4298	0.4386	0.4656
Lab	7	2	4	7	4	4	1	7	8	1
	0.4659	0.4723	0.4768	0.4790	0.4876	0.4914	0.5110	0.5148	0.5177	0.5246

Table 1: Depths of each $\hat{W}_{i,j}$, i.e. each estimated deviation from the laboratory–specific mean

matrix together with the mean and the median of the posterior distributions of $\lambda_k^{(1)}$, for $k = 1, \ldots, K_1$, for each one of the four different values of σ_λ^2. Our estimates for the $\lambda_k^{(1)}$'s are quite robust. Nevertheless, as the hyperparameter σ_λ^2 increases, for each k, the posterior mean and the posterior median slightly increases and so does the difference between the two, with a tendency to a positive skew (i.e. to a longer tail on the right) for the posterior

		1	2	3	4	5	6
	$\hat{\lambda}_k^{(1)}$	298.973	81.562	50.765	45.135	20.438	5.842
$\sigma_\lambda^2 = 10^1$	mean	299.035	81.662	50.911	45.622	21.727	6.278
	median	299.111	81.617	50.771	45.458	21.431	5.724
$\sigma_\lambda^2 = 10^2$	mean	299.035	81.662	50.911	45.622	21.727	6.278
	median	299.111	81.617	50.771	45.458	21.431	5.724
$\sigma_\lambda^2 = 10^3$	mean	300.158	88.190	56.114	60.513	32.012	6.526
	median	297.947	83.983	52.455	54.858	28.949	5.714
$\sigma_\lambda^2 = 10^4$	mean	309.208	92.583	58.752	66.386	32.054	6.590
	median	294.639	82.934	51.987	59.180	28.396	5.567

Table 2: Estimated level one eigenvalues

distribution of $\lambda_k^{(1)}$.

The different estimates obtained for level–two eigenvalues can be found in Table 3. Here, we have disagreement between our Bayesian estimates and the empirical estimates based on the within covariance matrix. Indeed, according to the Bayesian model, the biggest eigenvalue is the second one, regardless of the choice of the hyperparameters, while it is the first one according to the empirical estimates.

As final comment on this robustness analysis, we believe that the posterior estimates are quite stable and the prior setting does not substantially affect the statistical findings. We have reported in Section 6 the results under $\sigma_\lambda^2 = 10^3$, being this value an usual choice for vague priors.

8 Comparison of the point and functional approaches

As mentioned several times, one of the main goals of this analysis is the study of the variability of the observations taking into account the hierarcichal structure of the data: indeed, the true quantity of interest for the engineers are the repeatability and reproducibility errors (i.e. the between an within standard deviations). The main aspect of the analysis conducted here is to consider the data as functional while the mixed effect regression model is usually

		1	2	3	4	5
	$\hat{\lambda}_k^{(2)}$	20.251	18.425	1.481	1.147	0.267
$\sigma_\lambda^2 = 10^1$	mean	21.832	22.654	7.368	2.014	0.179
	median	21.520	22.380	7.015	1.535	0.139
$\sigma_\lambda^2 = 10^2$	mean	21.832	22.654	7.368	2.014	0.179
	median	21.520	22.380	7.015	1.535	0.139
$\sigma_\lambda^2 = 10^3$	mean	26.571	34.341	7.606	2.052	0.174
	median	25.487	32.713	7.233	1.572	0.135
$\sigma_\lambda^2 = 10^4$	mean	26.853	34.935	7.625	1.966	0.176
	median	25.838	33.602	7.263	1.485	0.136

Table 3: Estimated level–two eigenvalues

fitted pointwise (i.e. at each frequency band, see for instance Scamoni et al. 2009). Clearly, the pointwise analysis is quite simple, but some information is lost. In fact, in the pointwise case, the frequency dependency of the response variable $D_{2m,nT}(t)$ is discarded. To evaluate the gain by taking into account this dependency we estimate a Bayesian fixed effect model for each frequency. In particular, for each frequency t in $\{50, 63, \ldots, 5000\}$, we fit the following:

$$
\begin{aligned}
&Y_{ij}(t) = \mu_0(t) + \gamma_j(t) + \varepsilon_{ij}(t), \quad i = 1, \ldots, 8,\ j = 1, \ldots, 5 \\
&\gamma_1(t) \ldots \gamma_8(t) \mid \sigma_\gamma^2(t) \overset{iid}{\sim} N(0, \sigma_\gamma^2(t)) \\
&\varepsilon_{ij}(t) \mid \sigma_\varepsilon^2(t) \overset{iid}{\sim} N(0, \sigma_\varepsilon^2(t)) \\
&\mu_0(t) \sim N(0, 1000) \\
&\sigma_\gamma^2(t) \sim \text{inv–gamma}(2, 0.1), \quad \sigma_\varepsilon^2(t) \sim \text{inv–gamma}(2, 0.001).
\end{aligned}
$$

The prior choice for such models is quite standard and represents vague prior information (see Gelman et al. 2006).

The posterior pointwise estimation of the random effects γ's is reported in Table 4. The posterior credible bands under the functional analysis are smaller than the pointwise ones. Indeed, by taking into account the frequency dependence, we obtain a more precise posterior estimation.

Moreover, we can compare estimates of $\sigma^2_\gamma(t)$ with $\hat{G}_B(t,t)$ and $\sigma^2_\gamma(t) + \sigma^2_\varepsilon(t)$ with $\hat{G}_T(t,t)$. We do not report any pointwise estimate for brevity, while the functional counterparts are depicted in Fig. 2. We mention that estimates are similar almost on the whole frequency spectrum, apart from the 50 Hz frequency, where the pointwise estimates are much smaller than the functional ones. Consequently, in the pointwise analysis we observe a sharp increase of the estimated variances from 50 Hz to 63 Hz. This behaviour that was difficult to interpret by the engineers (see Scrosati et al. 2015) is mitigated in the functional framework thanks to the smoothing effect.

9 Conclusions

In this paper, we analyze data coming from a RRT experiment conducted at the ITC–CNR. Our aims were two: we wanted to identify those frequencies at which measurements are more variable and meanwhile to decompose the total variance according to the hierarchical nature of the data. An important achievement relies on the capability, under our modeling, to rank the measurement perfomances of the laboratories taking into account both the behaviour at low frequencies (high varibility) and the global perfomance. To rank laboratories according to their measurament correctness at the low frequencies we ground on the posterior of the scores of the first principal component at level one (i.e. laboratory level), while the global performance was evaluated according to the centrality of the estimated laboratory–specific deviation from the general mean, $\hat{Z}_i(t)$. A natural tool to measure the centrality of curves is the band depth: we briefly discussed the probabilistic setting of the MBD depth and use this tool to achieve our ranking goal.

In the recent literature, there are several papers dealing with multilevel regression functional models. Among the others, we have seen the model proposed by Di et al. (2009) fit to our aims. To make our exposition self–contained, we briefly reviewed the model and we have explained in details our estimation procedure. All the analysis was conducted using the R software interfacing it with JAGS when needed. Our model fits very well the data and it has proven an appropriate tool for our goals.

We found several benefits using a functional approach: the estimated functional correla-

tions among measuraments at different frequencies has provided additional insights into the data; the principal functional decomposition of the variance at both levels of the hierarchy matches the assumptions of the engineers on the functional directions of higher variability; the functional estimation of the random effects are more precise. In addition, the model is quite robust with respect to the hyperparameters of the scores priors.

From an acoustic point of view, the analyses illustrated in the paper led to the following findings. The first principal component, both for the first and the second level, reflects the low frequencies variability (from 50 to 100 Hz). Indeed, they both take high values (in absolute value) between 50 and 100 Hz, while they take values not far from zero at frequencies greater than 100 Hz. For each level, the first component explains a high percentage of the variability (55.23 % for level one and 48.54% for level two). We can also evaluate the proportion of the total variability due to the frequencies between 50 Hz and 100 Hz, by using the estimated intra–cluster correlation (88.5%). This can be determined at about 54%, derived as $55.23 \cdot 0.885 + 48.54 \cdot (1 - 0.885) = 0.5446$.

It was also shown that the best laboratory (i.e. the laboratory whose mean is the most central one on the basis of the MBD depth) is laboratory one, that is the laboratory where the rotating microphone boom was used for the measurement of the space and time averaged pressure level in the receiving room, while all the other laboratories used the fixed microphones positions.

Finally, it was found that a laboratory which tends to overestimate at the lowest frequencies will overestimate at the highest frequencies and vice versa.

References

Baladandayuthapani, V., Mallick, B., Young Hong, M., Lupton, J., Turner, N. & Carroll, R. (2008), 'Bayesian hierarchical spatially correlated functional data analysis with application to colon carcinogenesis', *Biometrics* **64**(1), 64–73.

Bigelow, J. & Dunson, D. (2007), 'Bayesian adaptive regression splines for hierarchical data', *Biometrics* **63**(3), 724–732.

Brumback, B. A. & Rice, J. A. (1998), 'Smoothing spline models for the analysis of nested and crossed samples of curves', *Journal of the American Statistical Association* **93**(443), 961–976.

Crainiceanu, C. & Goldsmith, A. (2010), 'Bayesian functional data analysis using WinBUGS', *Journal of Statistical Software* **32**(i11).

Crainiceanu, C., Staicu, A.-M. & Di, C.-Z. (2009), 'Generalized multilevel functional regression', *Journal of the American Statistical Association* **104**(488), 1550–1561.

Di, C.-Z., Crainiceanu, C. M., Caffo, B. S. & Punjabi, N. M. (2009), 'Multilevel functional principal component analysis', *Ann. Appl. Stat.* **3**, 458–488.

Fraiman, R. & Muniz, G. (2001), 'Trimmed means for functional data', *Test* **10**(2), 419–440.

Gelman, A. et al. (2006), 'Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper)', *Bayesian analysis* **1**(3), 515–534.

Guo, W. (2002), 'Functional mixed effects models', *Biometrics* **58**(1), 121–128.

Hall, P. & Hosseini-Nasab, M. (2006), 'On properties of functional principal components analysis', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **68**(1), 109–126.

Hall, P., Müller, H.-G. & Yao, F. (2008), 'Modeling sparse generalized longitudinal observations with latent Gaussian processes', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **70**, 703–723.

James, G., Hastie, T. & Sugar, C. (2000), 'Principal component models for sparse functional data', *Biometrika* **87**, 587–602.

Liu, R. Y., Parelius, J. M. & Singh, K. (1999), 'Multivariate analysis by data depth: descriptive statistics, graphics and inference, (with discussion and a rejoinder by Liu and Singh)', *The Annals of Statistics* **27**(3), 783–858.

Lopez-Pintado, S. & Romo, J. (2009), 'On the concept of depth for functional data', *J. Am. Statist. Assoc.* **104**, 718–734.

Morris, J. S. & Carroll, R. J. (2006), 'Wavelet-based functional mixed models', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **68**(2).

Morris, J. S., Vannucci, M., Brown, P. & Carroll, R. (2003), 'Wavelet-based nonparametric modeling of hierarchical functions in colon carcinogenesis', *Journal of the American Statistical Association* **98**(463), 573–583.

Plummer, M. et al. (2003), JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling, *in* 'Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003). March', pp. 20–22.

Ramsay, J. & Dalzell, C. (1991), 'Some tools for functional data analysis', *Journal of the Royal Statistical Society. Series B (Methodological)* pp. 539–572.

Ramsay, J. & Silverman, B. (2002), *Applied functional data analysis*, Springer, New York.

Ramsay, J. & Silverman, B. (2005), *Functional data analysis*, second edition edn, Springer, New York.

Scamoni, F., Scrosati, C., Mussin, M., Galbusera, E., Bassanino, M., Zambon, G. & Radaelli, S. (2009), Repeatability and reproducibility of field measurements in buildings, *in* 'Euronise09', Edinburgh, Scotland, UK.

Scrosati, C., Scamoni, F., Bassanino, M., Mussin, M. & Zambon, G. (2013), 'Uncertainty analysis by a round robin test of field measurements of sound insulation in buildings: Single numbers and low frequency bands evaluation - airborne sound insulation', *Noise Control Engr. J.* **61**(3), 291–306.

Scrosati, C., Scamoni, F. & Zambon, G. (2015), Uncertainty of façade sound insulation in buildings by a round robin test. submitted for publication.

Silverman, B. (1996), 'Smoothed functional principal components analysis by choice of norm', *The Annals of Statistics* **24**, 1–24.

Staicu, A.-M., Crainiceanu, C. & Carroll, R. (2010), 'Fast methods for spatially correlated multilevel functional data', *Biostatistics* **11**.

Sun, Y. & Genton, M. (2011), 'Functional boxplots', *J. Comput. Graph. Stat.* **20**, 316–334.

Uncertainty of measurement – Part 3: Guide to the expression of uncertainty in measurement (GUM:1995). International Standard ISO/IEC Guide 98-3: (2008). International Organization for Standardization, Geneva, Switzerland.

Yao, F., Müller, H.-G. & Wang, J. (2005), 'Functional data analysis for sparse longitudinal data', *Journal of the American Statistical Association* **100**(470), 577–590.

Freq.	Lab. 1						Lab. 2						Lab. 3					
	pointwise			functional			pointwise			functional			pointwise			functional		
	2.5 %	50 %	97.5 %	2.5 %	50 %	97.5 %	2.5 %	50 %	97.5 %	2.5 %	50 %	97.5 %	2.5 %	50 %	97.5 %	2.5 %	50 %	97.5 %
50	-4.382	-1.666	0.993	-4.558	-2.053	0.069	-4.630	-1.880	0.800	-5.034	-2.754	0.315	4.584	7.161	9.972	5.524	8.002	10.412
63	-6.096	-3.006	0.015	-4.813	-2.869	-1.046	-6.100	-2.959	0.079	-4.856	-3.144	-0.699	3.758	6.689	9.850	4.547	6.524	8.534
80	-4.614	-2.316	-0.125	-3.818	-2.446	-1.092	-2.228	-0.004	2.192	-2.205	-1.053	0.481	-0.482	1.607	3.850	1.262	2.566	3.866
100	-1.478	-0.419	0.574	-1.550	-0.789	-0.145	-0.258	0.749	1.862	-0.208	0.459	1.372	-0.254	0.738	1.817	0.052	0.690	1.417
125	-1.090	-0.189	0.557	-0.799	-0.049	0.652	-0.759	0.057	0.922	0.300	1.094	1.905	-0.421	0.298	1.263	-0.142	0.587	1.326
160	-0.724	0.958	2.652	-0.484	0.336	1.136	0.320	1.982	3.734	1.319	2.293	3.216	-0.680	0.965	2.651	-0.443	0.510	1.465
200	-1.207	-0.250	0.695	-0.643	0.069	0.741	1.020	1.954	2.969	1.607	2.405	3.216	-1.085	-0.144	0.791	-0.354	0.537	1.222
250	-1.038	0.318	1.630	-0.142	0.773	1.415	-0.134	1.200	2.550	0.263	1.042	1.741	0.896	2.204	3.548	0.612	1.319	2.095
315	0.882	1.960	3.067	1.487	2.111	2.816	-1.693	-0.598	0.491	-1.626	-0.772	-0.133	-0.167	0.894	1.976	1.014	1.744	2.409
400	0.964	2.171	3.364	1.236	1.898	2.670	-2.665	-1.435	-0.244	-1.940	-1.109	-0.326	1.231	2.419	3.619	1.207	1.862	2.594
500	-0.754	0.153	1.073	-0.100	0.683	1.413	-0.410	0.507	1.446	-0.128	0.680	1.475	0.089	0.971	1.909	0.193	1.014	1.808
630	-0.961	-0.000	0.944	-1.284	-0.391	0.419	1.770	2.705	3.732	2.149	2.951	3.731	-1.191	-0.246	0.690	-0.975	-0.093	0.670
800	-1.583	-0.615	0.313	-1.066	-0.507	0.071	1.336	2.274	3.257	1.775	2.406	3.059	-0.233	0.701	1.648	-0.571	0.219	0.852
1000	-0.303	0.742	1.763	-0.027	0.650	1.343	-1.231	-0.190	0.846	-0.846	-0.185	0.504	-0.442	0.582	1.611	0.295	0.864	1.520
1250	0.802	2.297	3.734	1.104	1.791	2.592	-3.710	-2.214	-0.779	-2.652	-1.868	-1.097	0.342	1.838	3.282	0.575	1.201	2.000
1600	0.011	0.404	0.810	0.742	1.078	1.397	0.116	0.512	0.928	-0.765	-0.301	0.068	0.053	0.439	0.842	0.428	0.842	1.182
2000	-0.072	0.872	1.778	0.150	0.777	1.229	-0.270	0.669	1.578	-0.107	0.502	1.135	0.084	1.029	1.942	0.135	0.941	1.373
2500	-0.083	0.780	1.679	-0.050	0.728	1.316	-1.181	-0.285	0.606	-0.007	0.640	1.316	0.770	1.616	2.531	0.427	1.374	1.918
3150	-0.540	0.434	1.375	-0.342	0.387	0.868	1.026	1.981	2.960	0.688	1.335	2.029	0.143	1.088	2.047	0.294	1.176	1.639
4000	-0.789	0.066	0.869	-0.570	-0.061	0.477	-1.541	-0.691	0.118	-0.567	-0.089	0.456	0.481	1.330	2.143	0.755	1.350	1.819
5000	-0.731	0.134	0.999	-1.059	-0.202	0.723	-1.563	-0.664	0.190	-2.024	-0.910	0.124	0.210	1.043	1.938	0.505	1.510	2.359

Freq.	Lab. 4						Lab. 5						Lab. 6					
	pointwise			functional			pointwise			functional			pointwise			functional		
	2.5 %	50 %	97.5 %	2.5 %	50 %	97.5 %	2.5 %	50 %	97.5 %	2.5 %	50 %	97.5 %	2.5 %	50 %	97.5 %	2.5 %	50 %	97.5 %
50	-0.491	2.108	4.804	-0.288	2.333	4.692	-6.028	-3.294	-0.656	-6.817	-3.651	-0.853	-1.233	1.410	4.117	-1.199	1.683	4.010
63	-0.886	2.070	5.107	0.681	2.793	4.763	-7.427	-4.316	-1.329	-7.106	-4.934	-3.003	-4.433	-1.391	1.615	-3.621	-1.468	0.488
80	-0.239	1.859	4.146	0.758	2.215	3.375	-5.177	-2.851	-0.627	-4.971	-3.434	-2.108	-4.462	-2.179	-0.005	-3.915	-2.487	-1.130
100	-0.421	0.557	1.624	0.338	1.119	1.769	-0.313	0.666	1.737	-0.981	-0.184	0.651	-1.262	-0.224	0.778	-1.168	-0.372	0.358
125	-0.651	0.108	0.974	0.247	1.140	1.735	-0.516	0.221	1.138	-0.408	0.292	1.047	-0.130	0.687	1.833	0.352	1.007	1.739
160	0.714	2.346	4.064	0.493	1.495	2.384	-2.945	-1.251	0.430	-1.897	-0.911	-0.011	-1.209	0.465	2.149	-0.649	0.275	1.200
200	-1.008	-0.067	0.871	-0.319	0.532	1.254	-1.990	-1.017	-0.082	-2.204	-1.388	-0.622	-2.140	-1.167	-0.239	-0.591	0.231	0.969
250	-0.673	0.646	1.943	-0.364	0.346	0.975	-2.561	-1.205	0.102	-1.212	-0.462	0.222	0.780	2.112	3.440	0.183	0.857	1.604
315	-0.767	0.297	1.363	0.291	0.926	1.709	0.237	1.299	2.405	-0.569	0.239	0.980	-3.174	-2.051	-1.005	-1.973	-1.277	-0.568
400	-0.151	1.044	2.220	-0.152	0.641	1.503	-2.623	-1.388	-0.216	-1.068	-0.335	0.476	-3.235	-1.994	-0.828	-3.637	-2.784	-1.986
500	-1.878	-0.946	-0.053	-1.328	-0.557	0.155	-1.093	-0.173	0.745	-1.402	-0.643	0.224	-2.787	-1.812	-0.915	-2.641	-1.846	-1.188
630	-2.340	-1.382	-0.472	-2.150	-1.343	-0.510	-1.667	-0.709	0.238	-1.878	-0.974	-0.149	-0.669	0.276	1.231	-1.083	-0.107	0.556
800	-1.015	-0.078	0.841	-0.577	0.224	0.884	-2.966	-1.993	-1.066	-2.369	-1.663	-1.032	-0.571	0.379	1.321	-0.805	-0.132	0.392
1000	1.104	2.114	3.140	1.622	2.194	2.724	-2.049	-0.990	0.023	-1.468	-1.004	-0.395	-2.472	-1.420	-0.418	-2.574	-2.075	-1.427
1250	1.075	2.570	4.003	1.864	2.559	3.270	-0.835	0.670	2.101	-0.693	-0.059	0.845	-4.723	-3.203	-1.772	-3.494	-2.748	-1.796
1600	-0.396	-0.002	0.390	-0.270	0.114	0.544	-1.307	-0.886	-0.498	-1.048	-0.689	-0.329	-0.160	0.235	0.636	-1.387	-1.023	-0.633
2000	-2.763	-1.810	-0.912	-1.926	-1.427	-0.849	-2.226	-1.270	-0.371	-1.882	-1.324	-0.705	-2.829	-1.868	-0.964	-1.821	-1.211	-0.654
2500	-1.674	-0.792	0.060	-2.155	-1.525	-0.798	-2.178	-1.269	-0.409	-2.189	-1.500	-0.831	-2.253	-1.347	-0.493	-2.190	-1.485	-0.912
3150	-3.009	-2.031	-1.116	-2.143	-1.519	-0.845	-2.652	-1.661	-0.727	-2.450	-1.792	-1.133	-1.811	-0.830	0.109	-1.701	-1.013	-0.420
4000	-0.308	0.540	1.336	-0.423	0.026	0.569	-2.796	-1.932	-1.134	-2.753	-2.212	-1.769	-2.086	-1.222	-0.424	-1.510	-1.073	-0.564
5000	-0.482	0.359	1.226	-0.379	0.465	1.420	-3.126	-2.182	-1.321	-3.433	-2.375	-1.540	-0.656	0.197	1.065	-0.789	-0.026	0.979

Freq.	Lab. 7						Lab. 8					
	pointwise			functional			pointwise			functional		
	2.5 %	50 %	97.5 %	2.5 %	50 %	97.5 %	2.5 %	50 %	97.5 %	2.5 %	50 %	97.5 %
50	-2.732	-0.048	2.620	-1.878	0.558	2.944	-6.282	-3.535	-0.947	-6.173	-3.680	-1.062
63	1.800	4.776	7.939	2.244	4.671	6.334	-4.591	-1.519	1.468	-3.365	-1.254	0.805
80	1.618	3.966	6.359	3.400	5.139	6.239	-2.078	0.054	2.213	-1.730	-0.368	0.920
100	-0.832	0.168	1.200	0.151	0.984	1.636	-3.493	-2.269	-0.935	-2.707	-1.875	-1.261
125	-0.985	-0.118	0.660	-1.752	-1.074	-0.338	-2.463	-1.146	-0.043	-3.698	-2.956	-2.181
160	-2.494	-0.793	0.866	-1.307	-0.369	0.462	-6.341	-4.560	-2.945	-4.667	-3.567	-2.588
200	0.774	1.700	2.686	-0.739	0.073	0.739	-1.935	-0.969	-0.050	-3.186	-2.327	-1.525
250	0.780	2.112	3.440	0.183	0.857	1.604	-3.846	-2.470	-1.176	-2.100	-1.427	-0.705
315	-3.174	-2.051	-1.005	-1.973	-1.277	-0.568	-0.737	0.348	1.425	-1.492	-0.792	-0.071
400	-3.235	-1.994	-0.828	-3.637	-2.784	-1.986	-0.569	0.651	1.829	0.317	0.982	1.800
500	-2.787	-1.812	-0.915	-2.641	-1.846	-1.188	0.903	1.803	2.769	0.605	1.319	2.042
630	-0.669	0.276	1.231	-1.083	-0.107	0.556	-0.484	0.464	1.416	-0.089	0.783	1.543
800	-0.571	0.379	1.321	-0.805	-0.132	0.392	-0.227	0.723	1.667	0.343	1.013	1.611
1000	-2.472	-1.420	-0.418	-2.574	-2.075	-1.427	0.440	1.473	2.503	0.921	1.558	2.264
1250	-4.723	-3.203	-1.772	-3.494	-2.748	-1.796	-1.050	0.466	1.909	0.173	0.920	1.695
1600	-0.160	0.235	0.636	-1.387	-1.023	-0.633	-0.742	-0.335	0.055	-0.230	0.195	0.566
2000	-2.829	-1.868	-0.964	-1.821	-1.211	-0.654	0.172	1.127	2.039	0.359	0.868	1.370
2500	-2.253	-1.347	-0.493	-2.190	-1.485	-0.912	0.305	1.162	2.061	0.553	1.196	1.808
3150	-1.811	-0.830	0.109	-1.701	-1.013	-0.420	-0.289	0.682	1.633	0.502	1.136	1.688
4000	-2.086	-1.222	-0.424	-1.510	-1.073	-0.564	0.949	1.805	2.621	0.986	1.608	2.031
5000	-0.656	0.197	1.065	-0.789	-0.026	0.979	-0.577	0.271	1.128	-0.393	0.749	1.603

Table 4: Comparison of pointwise and functional posterior quantiles for the effect of each laboratory