MOX-Report No. 31/2015

Domain-selective functional ANOVA for process analysis via signal data: the remote monitoring in laser welding

Pini, A.; Vantini, S.; Colombo, D.; Colosimo, B. M.; Previtali, B.

MOX, Dipartimento di Matematica
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

mox-dmat@polimi.it http://mox.polimi.it

Domain-Selective Functional ANOVA for Process Analysis via Signal Data: the Remote Monitoring in Laser Welding

A. Pini[a], S. Vantini[a], D. Colombo[b], B. M. Colosimo[b], B. Previtali[b]

[a] MOX– Department of Mathematics, Politecnico di Milano, Milan Italy
alessia.pini@polimi.it
simone.vantini@polimi.it

[b] Department of Mechanical Engineering, Politecnico di Milano, Milan Italy
daniele.colombo@polimi.it
biancamaria.colosimo@polimi.it
barbara.previtali@polimi.it

Abstract

In many application domains, process monitoring and process optimization have to deal with functional responses, also known as profile data. In these scenarios, a relevant industrial problem consists in discovering which specific parts of the functional response is mostly affected by the process changes. As a matter of fact, knowledge of the specific locations where the curve is more sensitive to process changes can bring several advantages. It can be exploited to design specific monitoring devices directly focusing on the functional data pertaining to the selected intervals. Secondly, the dimensional reduction can eventually bring to an increase of the power to detect process changes.

This paper proposes a methodology to inferentially select the parts of the output functions that are more informative in terms of the underlying factors. The procedure is based on a non-parametric domain-selective ANOVA for functional data, which results in the selection of the intervals of the domain presenting statistically significant effects of each factor. To illustrate its potential in industrial applications, the proposed procedure is applied to a case study on remote laser welding, where the main aim is monitoring the gap between the welded plates through the observation of the emission spectra of the welded material.
Keywords: Statistical Process Control, Design of Experiments, Functional Data Analysis, Family-Wise Error Rate

1 Introduction

Due to the advent of in-line sensoring and non-contact measurement systems, more and more often responses of manufacturing processes can be modeled as functional data, i.e., a vector of data, where each value corresponds to a specific temporal or spatial location. Functional data can in fact represent the shape of a machined feature and/or the pattern of a signal acquired by the process over time. In this scenario, statistical quality monitoring and process optimization have to be appropriately rethought to deal with functional responses. With reference to statistical quality monitoring, the problem of checking the stability of a functional response has been named profile monitoring and has attracted the attention of many researchers over the past ten years (for an overview, see, for instance, Noorossana et al. 2012 and references therein). In the field of process optimization, robust optimization of functional data received the attention of researchers in many different application domains (Nair et al. 2002; Del Castillo and Colosimo 2011; Alshraideh and Del Castillo 2014; Del Castillo et al. 2012; He et al. 2014; Drignei 2010).

Both process monitoring and optimization of functional responses share a common basic problem, i.e., to detect a change of the functional response pattern. In the first case, the change is deemed to be due to an out-of-control state of the manufacturing process. In the case of process optimization, the functional response change is pursued by changing the process parameters, with the aim of making the response following as close as possible a target pattern. Within such framework, the analysis of the whole response function can provide lots of information on the underlying interest factors. Nevertheless, the functional observation and analysis present some issues. For instance, the in-process monitoring of the whole output function can be computationally not affordable, and/or the effect of the factors on the whole function can be difficult to study. Hence, it is often required to reduce the dimensionality of the output data, to provide a methodology that is efficient and actually applicable in industrial environments.

Functional principal component analysis (Ramsay and Silverman 2005) can be a way to reduce the dimensionality, and to select the most informative features of the data in terms of the proportion of explained variance (i.e., the functional principal components). This method can be applied to problems dealing with profile data, to reduce the dimensionality of the output functions and better understand their characteristics and evolution during the industrial process (Hall et al. 2001; Colosimo and Pacella 2007, 2010; Yu et al. 2012; Grasso et al. 2014). Unfortunately, functional principal components are often difficult to interpret, and in the field of functional ANOVA, this analysis can be misleading, as the components of the signal explaining the higher variability are not necessarily the ones that present more information about the factors of interest of the industrial process. Furthermore, functional principal component analysis relies on the evaluation of all of the original functional data. This means that the entire data domain has to be acquired to detect a process change.

In many industrial applications, the output functions may be influenced by the factors only in some parts of their domain, or in different ways over the domain. To efficiently monitor the process, it is then key to select the informative parts of the functions (i.e. the ones influenced by the factors). The information on which parts of the domain are informative can be used to design specific monitoring devices that only focus on the data pertaining to those parts. This would provide a direct gain, both in terms of efficiency of the monitoring procedure, and in economic terms, since the monitoring would then be based on the univariate output of the sensor, instead of on more complicated evaluations based on the whole functions. As an example, the motivating case study of this research paper is remote laser welding, where the laser emission spectrum is used as reference to check whether the distance between the plates to be welded (i.e., the gap) is appropriate. In this case, a previous study Colombo et al. (2013) showed that the gap between the plates affects the emission spectra and hence this last information can be used for monitoring purposes. However, developing a specific sensor aimed at detecting the emission just at some specific wavelength instead of using the whole spectra can reduce costs of the monitoring device and reduce the computational time.

Even though the selection of the informative parts of the domain can be done in naive ways, for instance by a visual inspection of the data, or based on prior knowledge about the process, in most cases visual inspection and/or prior knowledge can be unclear or misleading, and may neglect important aspects. Hence, we here propose a statistical technique to approach the domain selection. The approach that we propose is sound even in absence of prior knowledge about the process, and gives clear and reliable results. We propose to select the informative parts of the functions by means of the inferential interval-wise testing procedure, first proposed in Pini and Vantini (2015), and here extended to the case of multi-way functional ANOVA. In detail, we test the significance of the effects of the factors on the output functions. More importantly, we provide a selection of the parts of the domain presenting significant effects of each factor. Once this selection has been done, we propose to monitor the industrial process based on the information that the functions contain on the selected domains.

The paper is structured as follows: Section 2 presents the statistical methodology applied for the domain-selective functional ANOVA. Section 3 reports an application of the proposed method on a remote laser welding data set. In detail, Subsection 3.1 reports the description of the industrial application and the related problems, Subsection 3.2 reports the description of the data, and Subsection 3.3 reports the results of the domain-selective functional ANOVA on these data. A robustness analysis is reported in the Appendix.

2 Domain-selective functional ANOVA

In this section we describe the methodology that we apply to test a multi-way functional ANOVA. For ease of notation, and without loss of generality, we consider the example of a two-way functional ANOVA with interaction. Nevertheless, the methodology presented here can be directly generalized to the analysis of factorial design with more than two factors.

Consider a two-way ANOVA in which we investigate the effects of factors A and B (with I and J levels, respectively), on a functional response, based on a replicated full factorial design. We assume that the response of the model is a function of a continuous variable t observed over a domain $(a, b) \subset \mathbb{R}$. In the considered model, the functional response will be expressed as the result of two main effects, and of an interaction between the previous ones. The aim of the analysis is to test the significance of every term in the model (i.e., interaction and main effects).

In detail, let $y_{ijl} : (a, b) \mapsto \mathbb{R}$ be the output functional data, where i denotes the level of the first factor, j denotes the level of the second factor, and l the replicate. We here assume functional data to be continuous functions of the variable t. Note that, as shown in Pini and Vantini (2015), this condition can be relaxed, and the methodology can be applied to generic L^2 functions. The functional ANOVA model that we want to test is thus the following:

$$y_{ijl}(t) = \mu(t) + \alpha_i(t) + \beta_j(t) + \gamma_{ij}(t) + \epsilon_{ijl}(t), \tag{1}$$

where $t \in (a, b)$, with $i = 1, \dots, I$, $j = 1, \dots, J$, $l = 1, \dots, n_{ij}$, n_{ij} being the number of replicates for ith level of the first factor and jth level of the second factor. In model (1), $\mu(t)$ is the functional grand mean, $\alpha_i(t)$ and $\beta_j(t)$ are functional main effects, and $\gamma_{ij}(t)$ is the functional interaction effect. The functional errors $\epsilon_{ijl}(t)$ are assumed to be independent and identically distributed zero-mean random functions. Note that we do not require the errors to follow a gaussian process. All effects (as well as the errors) are expressed as functions of the continuous variable t. For sake of identifiability, we require the classical constraints on the effects, i.e., $\forall t \in (a, b)$: $\sum_{i=1}^{I} n_i \alpha_i(t) = 0$; $\sum_{j=1}^{J} n_j \beta_j(t) = 0$; $\sum_{i=1}^{I} \sum_{j=1}^{J} n_{ij} \gamma_{ij}(t) = 0$, where $n_i = \sum_{j=1}^{J} n_{ij}$ denotes the number of units at ith level of the first factor and $n_j = \sum_{i=1}^{I} n_{ij}$ denotes the number of units at jth level of the second factor.

The aim of our analysis is to test the significance of all coefficients of model (1). In particular, we want to perform the functional counterparts of uni/multivariate ANOVA tests:

- a functional test of the null model, jointly for all factors:

$$\begin{cases} H_{0,Model} & : \alpha_i(t) = \beta_j(t) = \gamma_{ij}(t) = 0 \ \forall i = 1, \dots, I; j = 1, \dots, J; t \in (a, b); \\ H_{1,Model} & : (H_{0,Model})^C; \end{cases} \tag{2}$$

- three functional tests for the effects of each factor and interaction:

$$H_{0,A}: \ \alpha_i(t) = 0 \ \forall i = 1, \ldots, I; t \in (a, b); \quad H_{1,A} : (H_{0,A})^C \tag{3}$$

$$H_{0,B}: \ \beta_j(t) = 0 \ \forall j = 1, \ldots, J; t \in (a, b); \quad H_{1,B} : (H_{0,B})^C \tag{4}$$

$$H_{0,AB}: \ \gamma_{ij}(t) = 0 \ \forall i = 1, \ldots, I; j = 1, \ldots, J; t \in (a, b); \quad H_{1,AB} : (H_{0,AB})^C. \tag{5}$$

Note that, similarly to uni/multivariate two-way ANOVA, (2) is a test of significance of the whole functional model, whereas the three tests (3-5) allow to perform a model selection. The main difference with respect to a uni/multivariate ANOVA is that, being the response functional, tests (2-5) involve functional coefficients, i.e., the null hypothesis is rejected whenever there is a significant difference between the corresponding groups in at least one interval of the domain.

The problem of testing a functional ANOVA model has been widely discussed in the literature of the last decades, and it can be addressed in several ways (e.g.,Cuevas et al. (2004); Cuesta-Albertos and Febrero-Bande (2010); Abramovich and Angelini (2006); Antoniadis and Sapatinas (2007)). A common feature of all these works, is that the final result from the ANOVA testing determines whether the hypotheses (2-5) are globally accepted or rejected. In particular, by applying these tests, we only are able to answer the question "Are there any statistically significant effects of factors A and/or B on the functional responses?". In the case of a positive answer, these tests are not able to select the intervals of the domain in which the effects are detected.

On the contrary, the principal aim of this paper is to provide practitioners with a methodology that selects the most informative parts of the functions domain, in order to use this information to design specific monitoring devices. In detail, for each tests (2-5), in case of rejection of the null hypothesis, we want to select the intervals of the variable t where significant differences are detected. For this reason, we extend the interval-wise testing proposed in (Pini and Vantini 2015), that is a testing procedure for functional data that enables to select the intervals of the domain presenting significant effects.

Another advantage of the application of such procedure is being a non-parametric procedure. In particular, we neither need to assume the normality of the residuals of the model, nor to specify their covariance structure, that can be both difficult to assess in the practice.

The extension of the procedure on functional ANOVA is based on three steps, detailed in the following paragraphs.

First step: interval-wise testing. A functional test corresponding to tests (2-5) is performed on any interval of the domain. In detail, given any $\mathcal{I} \in (a, b)$, we perform a test of the null model, and three tests on the main effects and interaction:

$$\begin{cases} H^{\mathcal{I}}_{0,Model} & : \alpha_i(t) = \beta_j(t) = \gamma_{ij}(t) = 0 \quad \forall i = 1, \ldots, I; j = 1, \ldots, J; t \in \mathcal{I}; \\ H^{\mathcal{I}}_{1,Model} & : (H^{\mathcal{I}}_{0,Model})^C \end{cases} \tag{6}$$

$$H^{\mathcal{I}}_{0,A} : \ \alpha_i(t) = 0 \ \forall i = 1, \ldots, I; t \in \mathcal{I}; \quad H^{\mathcal{I}}_{1,A} : (H^{\mathcal{I}}_{0,A})^C \tag{7}$$

$$H^{\mathcal{I}}_{0,B} : \ \beta_j(t) = 0 \ \forall j = 1, \ldots, J; t \in \mathcal{I}; \quad H_{1,B} : (H^{\mathcal{I}}_{0,B})^C \tag{8}$$

$$H^{\mathcal{I}}_{0,AB} : \ \gamma_{ij}(t) = 0 \ \forall i = 1, \ldots, I; j = 1, \ldots, J; t \in \mathcal{I}; \quad H^{\mathcal{I}}_{1,AB} : (H^{\mathcal{I}}_{0,AB})^C. \tag{9}$$

As proposed in Pini and Vantini (2015), we perform each test in a permutation framework. In detail, we apply permutations of the residuals of the reduced model, according to the Freedman and Lane permutation scheme (Freedman and Lane 1983). As test statistics, we compute the integral over the interval $\mathcal{I}$ of the two-way ANOVA statistics of the corresponding classical F-tests. This provides exact tests for $H_{0,Model}$, and approximated (asymptotically exact) tests for $H_{0,A}$, $H_{0,B}$ and $H_{0,AB}$. In the following, we denote with $p^{\mathcal{I}}_{Model}$, $p^{\mathcal{I}}_{A}$, $p^{\mathcal{I}}_{B}$, and $p^{\mathcal{I}}_{AB}$ the p-values of tests (6-9), respectively.

Second step: computation of the adjusted p-value function. An adjusted p-value function is computed for each test (2-5). In detail, for each $t \in (a, b)$, the p-value $\tilde{p}(t)$ is computed as the maximum p-value of all interval-wise tests (6-9) on intervals containing t:

$$\tilde{p}_{Model}(t) = \sup_{\mathcal{I} \ni t} p^{\mathcal{I}}_{Model}, \quad \tilde{p}_A(t) = \sup_{\mathcal{I} \ni t} p^{\mathcal{I}}_{A}, \quad \tilde{p}_B(t) = \sup_{\mathcal{I} \ni t} p^{\mathcal{I}}_{B}, \quad \tilde{p}_{AB}(t) = \sup_{\mathcal{I} \ni t} p^{\mathcal{I}}_{AB}.$$

The adjusted p-value function $\tilde{p}_{Model}(t)$ is provided by a control of the interval-wise error rate, while the adjusted p-value functions $\tilde{p}_A(t)$, $\tilde{p}_B(t)$, and $\tilde{p}_{AB}(t)$ are provided with an asymptotic control of the interval-wise error rate, as defined in Pini and Vantini (2015).

Third step: domain selection. The intervals of the domain presenting a rejection of the null hypothesis are obtained by thresholding the corresponding adjusted p-value function at level α: we select intervals presenting at least one significant effect by thresholding $\tilde{p}_{Model}(t)$, intervals presenting a significant effect of the A factor by thresholding $\tilde{p}_A(t)$, and so on.

The control of the interval-wise error rate allows controlling the probability of detecting false positive intervals, i.e., the probability of wrongly rejecting any interval. For instance, since $\tilde{p}_{Model}(t)$ is provided by a control of the interval-wise error rate, by selecting the intervals associated with an adjusted p-value $\tilde{p}_{Model}(t) \leq \alpha$, we have that, given any interval in which the response is not influenced by any factor, the probability that this interval is (wrongly) selected as significant is lower than α.

3 Case study: remote monitoring of remote laser welding

To illustrate the potential of the approach proposed in this paper in industrial applications, in the following we report an application on remote laser welding data. During the remote welding of zinc-coated steel in the lap-joint configuration, we register as output profile the optical emission of the welded material as a function of the wavelength. We are interested in selecting the informative part of these profiles for the monitoring of the gap between the two welded surfaces, by taking into account the location in which the emission spectra are acquired. In the following, we report the description of the industrial process of remote laser welding, and the related issues (Subsection 3.1), the description of the data (Subsection 3.2), and the results of the analysis (Subsection 3.3). A robustness study with respect to the smoothing parameter is reported in the Appendix.

3.1 Remote monitoring of laser welding

Laser welding technologies are quickly replacing conventional welding processes. Furthermore, nowadays the laser welding is often used in remote configurations, i.e., configurations in which the laser beam is moved along the seam with the help of a laser scanner. One of the most common applications of such process is the welding of zinc-coated steel in the lap-joint configuration. However, this particular configuration and materials present a lot of technical issues, since the boiling temperature of the zinc is significantly lower than the one of steel (approx. 906°C and 1500°C, respectively). Consequently, during the welding, highly pressurized zinc vapors are often produced at the interface of the two metal sheets, and may cause defects in the welded material, such as spatters and porosities, that can compromise its quality.

The classical solution applied in industrial processes to prevent these defects is to leave a small gap (order of hundreds microns) between the two metal sheets, to facilitate the degassing (Akhter et al. 1991; Steen et al. 2003). One of the methods used to produce such gap is laser dimpling, which uses a remote pulsed laser to generate protuberances on one of the plates (Daimler Chrysler AG 2005; Schwoerer 2008; Gu 2010). However, the variance of the height of laser-dimples can cause errors in the final gap dimension, which can cause defects on the welded material, compromising its quality and causing variations in the mechanical properties of the weld bead.

In Colombo et al. (2013), a method to remotely monitor the gap in remote laser welding, avoiding destructive off-line tests, is proposed. According to this technique, optical emissions are monitored during remote laser welding by a spectroscope. Then, from the acquired spectra, different indicators, or summarizing variables, are evaluated, such as the overall emission across the considered range, and the emissions in separated wavelength ranges (defined by physical evaluations of the welding process). For each of the obtained variables, univariate analysis of variance is performed, and the statistical significance of the effects of the gap value is used to compare the tested methods. The optical emission recorded during the experiment can also depend on the location in the weld seam where the spectrum is acquired. That is why, in Colombo et al. (2013), a two-way analysis of variance is performed, to evaluate the effect of the gap on the emission taking into account the different locations. The former analysis is a valid instrument to assess how the emission is influenced by the gap, and the results can be used to provide an indicator to evaluate the gap remotely. However, the choice of the better indicator to evaluate the gap effect is difficult to perform, since the analyzed wavelength bands are priorly fixed.

Here, we perform a domain selective functional ANOVA, according to the procedure described in Section 2, to select the wavelength bands presenting significant effects of the gap and the location on the emission, and controlling the probability of false discoveries. The direct result of this analysis is the selection of the wavelength bands that are informative in terms of the gap between the plates. The selected bands can thus be used to remotely monitor the gap between the plates during the welding process.

3.2 Experimental procedure and data acquisition

A Through Optical Combiner Monitoring architecture (Capello et al. 2008; Colombo and Previtali 2009, 2010) is used to monitor the laser welding. According to this technique, the monitoring of laser emission is performed remotely. Indeed, far from the work area, the optical emissions from the welding process are directly observed inside the laser source through the optical combiner of the fiber laser source with a spectroscope. An extensive description of the experimental welding procedure and the monitoring technology goes beyond the scope of this paper, and can be found in Colombo et al. (2013).

The main objective of this study is to assess the effects of both gap and location on the emission functions. To analyze these effects, the emission is acquired in correspondence of different levels of gap and location, in a repeated factorial design. Three values of gap, corresponding to 100 nm, 200 nm and 300 nm are explored. For each of the analyzed gap values, three replicates are produced, for a total of nine welded specimens. Inside each specimen, five emission spectra are acquired at five different locations, for a total number of 45 acquired spectra. The emission data are described in detail in Colombo et al. (2013).

To record the optical emission in the visible range, optical emission spectroscopy, i.e., the analysis of emitted light with high-wavelength resolution, is used. The laser emission $y_{ijl}(t)$ is acquired at 703 discrete wavelengths t between 400.521 nm and 800.030 nm (where indexes i, j, l indicate the levels of gap, location and replicate, respectively). The acquired data, as well as the three means, corresponding to the three different values of the gap, are represented in Figure 1. In the figure, the curves of each color correspond to the 15 functional

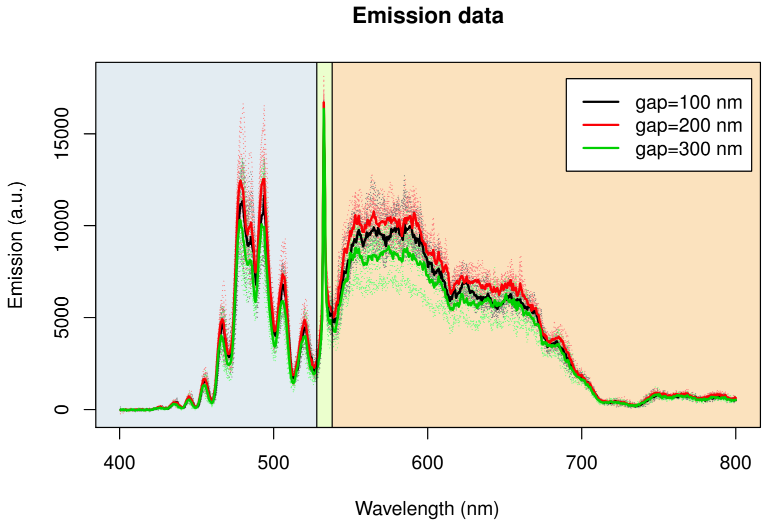

Figure 1: Emission data captured in the factorial laser welding experiment (dashed lines), colored according to the three different gaps; and means of the three groups, corresponding to the three different gaps (solid lines). The figure is divided along the abscissa into the plasma (light blue), laser (light green), and thermal (light red) emissions.

emissions (five locations and three replicates) within each gap level.

Note that in the explored wavelength range, it is possible to distinguish between three different emission ranges:

- between 400 nm and 530 nm, it is observed the emission related to electronic transition, i.e., the plasma emission (light blue area of Figure 1);
- around 535 nm we observe a strong emission line, corresponding to the laser emission (light green area of Figure 1);
- above 540 nm we observe the emission due to the thermal black-body radiation, i.e. the thermal emission (light red area of Figure 1).

The aim of the following analysis is to assess whether the gap and the location have some effects on the emission, by taking into account the whole functional shape of the spectrum, controlling the probability of false discoveries. Finally, we want to locate possible wavelength bands in which the significant effects are detected.

3.3 Results of the tests

In the application, the emission functions $y_{ijl}(t)$ ($i = 1, \ldots, 3$, $j = 1, \ldots, 5$, $l = 1, \ldots, 3$) are modeled according to model (1), where t is the wavelength, $\alpha_i(t)$ is the gap functional effect, $\beta_j(t)$ is the location functional effect, $\gamma_{ij}(t)$ is the interaction functional effect, and $\epsilon_{ijl}(t)$ is the functional error term.

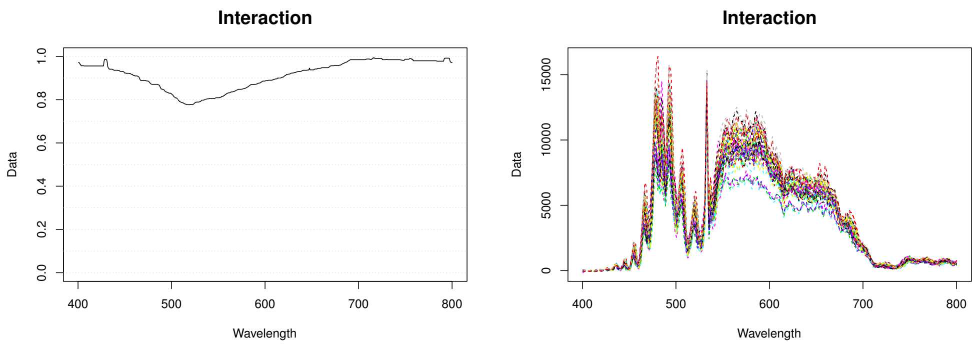

Figure 2: Left: adjusted p-value function for the tests on $\tilde{H}_{0,GapLoc}$. Right: emission data colored according to gap and location levels.

We apply the domain-selective functional two-way ANOVA described in Section 2 to test of the null model $H_{0,Model}$ (2), the gap effect $H_{0,Gap} = H_{0,A}$ (3), the location effect $H_{0,Loc} = H_{0,B}$ (4), and the interaction effect $H_{0,GapLoc} = H_{0,AB}$ (5). The functional data that we analyze are based on a smoothing of the 703 discrete data points on a dense piece-wise linear B-spline basis characterized by 200 knots. The obtained data after the smoothing are reported in the right panels of Figure 3. A robustness analysis with respect to the number of knots of the expansion is reported in the Appendix. The test statistics of tests on intervals were approximated through a rectangle integration method. Thanks to the continuity of test statistics with respect to the extremes of integration, the adjusted p-value functions were approximated through a finite family of tests over a fine discrete grid of 703 points (i.e., the maximal resolution of the monitoring instrument). The permutation p-values were estimated by means of a Conditional Monte Carlo algorithm based on 1000 iterations.

We started by applying the testing methodology described in Section 2 to test the interaction term $\gamma_{ij}(t)$. The results of the test are reported in Figure 2. In detail, the left panel reports the adjusted p-value function for the test of $H_{0,GapLoc}$, and the right panel reports the emission data obtained after the B-spline smoothing. Since the interaction effect is not significant (i.e., the adjusted p-value function presents high values along all the wavelength domain), we decided removing this term from the model, and tested the following additive functional ANOVA model:

$$y_{ijl}(t) = \mu(t) + \alpha_i(t) + \beta_j(t) + \epsilon_{ijl}(t), \quad \forall t \in (a, b). \tag{10}$$

In the following, we describe in detail the results of the tests on the additive model (10). Note that, in this case, the test of the null model becomes:

$$\begin{cases} \tilde{H}_{0,Model} & : \ \alpha_i(t) = \beta_j(t) = 0, \ \forall i = 1, \ldots, 3; j = 1, \ldots, 5; t \in (400.521, 800.030); \\ \tilde{H}_{1,Model} & : (\tilde{H}_{0,Model})^C. \end{cases}$$

Figure 3 reports the results of the additive two-way functional ANOVA of emission data. In particular, the left panels report the adjusted p-value functions for each test. In detail, on the top panel, we report the adjusted p-value functions of $\tilde{H}_{0,Model}$, on the middle panel the adjusted p-value functions of $H_{0,Gap}$ and on the bottom panel the adjusted p-value functions of $H_{0,Loc}$. For ease of visualization of the test results, the right panels of the figure report the emission data and the significant intervals detected at 5% and 1% levels (areas colored in light and dark grey, respectively) for the three tests. Data are colored in the three panels according to the corresponding tests: on the top panel data are colored differently according to the different levels of both gap and location; on the middle panel data are colored according to the different levels of gap; on the bottom panel data are colored according to the different levels of location.

Focusing on the test of $\tilde{H}_{0,Glob}$ (i.e., top panels of Figure 3), we find a significant effect of at least one factor among gap and location in nearly the entire wavelength domain. This result suggests that, as expected, the welding conditions have a significant effect on the spectrum for all three emission ranges (plasma, laser, and thermal). The gap has a significant effect in most of the wavelength domain (i.e., test of $H_{0,Gap}$, middle panels of Figure 3), suggesting that, consistently with previous results, the emission is significantly influenced by the gap on all three emission ranges. The location has a significant effect mostly in the plasma emission range (i.e., test of $H_{0,Loc}$, bottom panels of Figure 3). This suggests that plasma emission is influenced by the location.

The procedure described in this paper improves the results found by Colombo et al. (2013), by precisely locating the wavelength bands associated to significant effects of the two factors. Indeed, the most important result highlighted by this analysis, and completely new with respect to the literature in this field, is that, looking at all tests together, we detect a band (i.e., the band $t \in$ (547 nm, 681 nm) corresponding approximately to the thermal emission), in which the gap effect is significant and the location one is not. This suggests the use of emission data on this band to monitor the gap between the plates during the welding process at any possible location.

Note that, for instance, the adjusted p-value function of the test of $H_{0,Glob}$ is lower than 5% on the interval (547 nm, 681 nm). Thanks to the control of the interval-wise error rate, if the emission were effected by neither the gap nor the location in such band, we would have selected it as significant with a probability lower that 5%. Furthermore, since the adjusted p-value function of the test of $H_{0,Gap}$ is lower than 1% on the interval (547 nm, 681 nm), thanks to the asymptotic control of the interval-wise error rate, if the gap were not affect the emission in such band, we would have selected it as significant with a probability approximately lower that 1%.

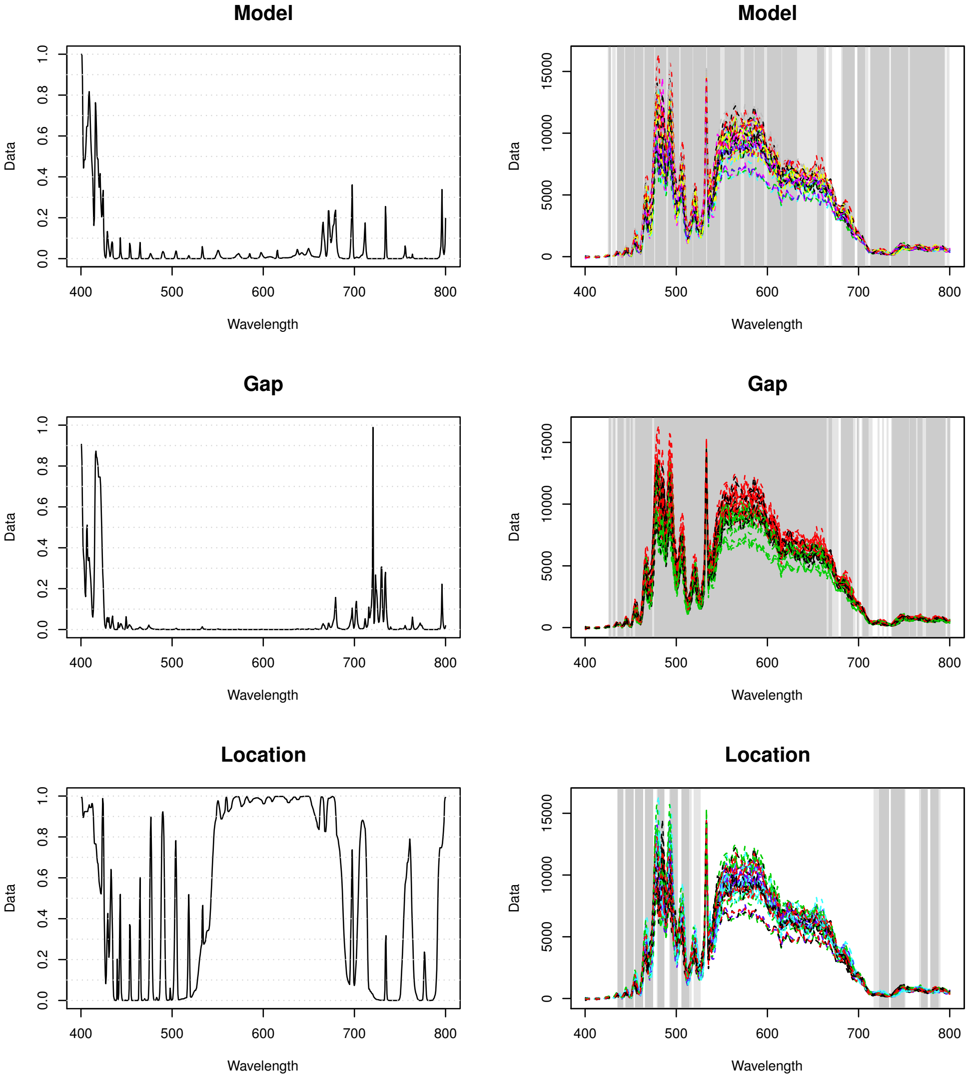

Figure 3: Left: adjusted p-value functions for the tests on $\tilde{H}_{0,Model}$ (top), $H_{0,Gap}$ (middle) and $H_{0,Loc}$ (bottom). Right: emission data colored according to gap and location levels (top), gap levels (middle) and location levels (bottom). The gray areas represent significant intervals at 5% and 1% levels (light and dark grey, respectively)

4 Conclusions

We proposed a methodology that can be applied in profile monitoring and functional response optimization to select the informative parts of the output functions. The methodology exploits a domain-selective functional ANOVA to study the effects of one or more underlying factors on the output functions. This selection is performed by extending the interval-wise testing procedure (Pini and Vantini 2015) to the case of multi-way ANOVA framework. This methodology represents a significant improvement with respect to the state of the art techniques. Indeed, it is a functional technique, in the sense that considers the whole functional data instead of performing a prior dimensional reduction. Furthermore, it selects the informative parts of the output functional data, in order to optimize the monitoring of the input factors through the observation of the output curves.

The proposed methodology is applied to a remote laser welding case study. The aim of this study is to remotely monitor the gap between two welded plates, by observing the emission spectra of the welded material. We applied the domain-selective functional two-way ANOVA to study the effects of the gap on the output functions, by taking into account the different locations at which the emission is registered.

The final result of this analysis is the selection of a wavelength band (i.e., the band $t \in$ (547 nm, 681 nm) corresponding to the thermal emission), in which the emission is significantly influenced by the gap and not by the location. This suggests to directly record the emission data on this band, to monitor the gap between the plates during the welding process at any possible location. The interval-wise control, that this procedure is based on, gives a direct measure of the reliability of this result. In detail, we know that, if the emission were effected by neither the gap nor the location in the band $t \in$ (547 nm, 681 nm), we would have selected it as significant with a probability lower that 5%.

The robustness of this result with respect to the number of knots p of the B-spline basis expansion employed to smooth the data is discussed in the Appendix. Even though the results of this robustness study show that in this case, the number of knots is not a crucial parameter, it is of course of major interest for future research to explore the trade-off, in terms of power of the procedure, as p increases.

An interesting future challenge posed by this data analysis is to introduce the gap in the functional model as a numeric explanatory variable, instead of a factor. Since the effect of the gap on the emission spectra is observed to be not monotonic, to perform this analysis, it is key to study its effect, and add the gap in the model in a non-linear way.

Appendix. Robustness analysis

To investigate the robustness of the results with respect to the number of knots of the B-spline expansion, we performed the test by varying this parameter. We considered different cases, based on B-splines expansions with a different number of knots. In particular, we started from the analysis reported in Section 3 based on 200 knots, and tried to double or halve the number of knots. Finally, we also tried the maximum possible resolution (i.e., 703 knots).

The test on the interaction $H_{0,GapLoc}$ is not significant in all cases. Hence, the comparison is made on the additive model without interaction (10). The results of the tests are reported in Figure 4. Similarly to the right panels of Figure 3, the results of the tests of $\tilde{H}_{0,Model}$, $H_{0,Gap}$ and $H_{0,Loc}$ at 1% and 5% levels are represented by means of the gray bands below each graphic. The axis on the left indicates the number of knots used for the B-spline basis expansion, i.e., 100, 200, 400, and 703 knots, corresponding to length of B-spline supports of 8, 4, 2, and 1.14 nm.

This further analysis shows that the results are robust with respect to the number of knots. Both the test on the null model $\tilde{H}_{0,Model}$ and the test on the gap $H_{0,Gap}$ report significant differences in most of the wavelength domain. In particular, the gap has a significant effect on all three emission ranges. On the other hand, the effect of the location is significant mostly on the plasma emission. In all cases, the thermal band detected in Subsection 3.3, presenting a significant effect of the gap but not of the location, is preserved.

References

Abramovich, F. and Angelini, C. (2006), Testing in mixed-effects fanova models, *J. Statist. Plann. Inference* **136**(12), 4326–4348.

Akhter, R., Steen, W. M. and Watkins, K. G. (1991), Welding zinc-coated steel with a laser and the properties of the weldment, *J. Laser Appl.* **3**(2), 9–20.

Alshraideh, H. and Del Castillo, E. (2014), Gaussian process modeling and optimization of profile response experiments, *Quality and Reliability Engineering International* **30**(4), 449–462.

Antoniadis, A. and Sapatinas, T. (2007), Estimation and inference in functional mixed-effects models, *Comput. Statist. Data Anal.* **51**(10), 4793–4813.

Capello, E., Colombo, D. and Previtali, B. (2008), Monitoring through the optical combiner in fibre laser welding, *in* 'ICALEO 2008 Congress proceedings', pp. 75–84.

Colombo, D., Colosimo, B. M. and Previtali, B. (2013), Comparison of methods for data analysis in the remote monitoring of remote laser welding, *Opt. Laser. Eng.* **51**(1), 34–46.

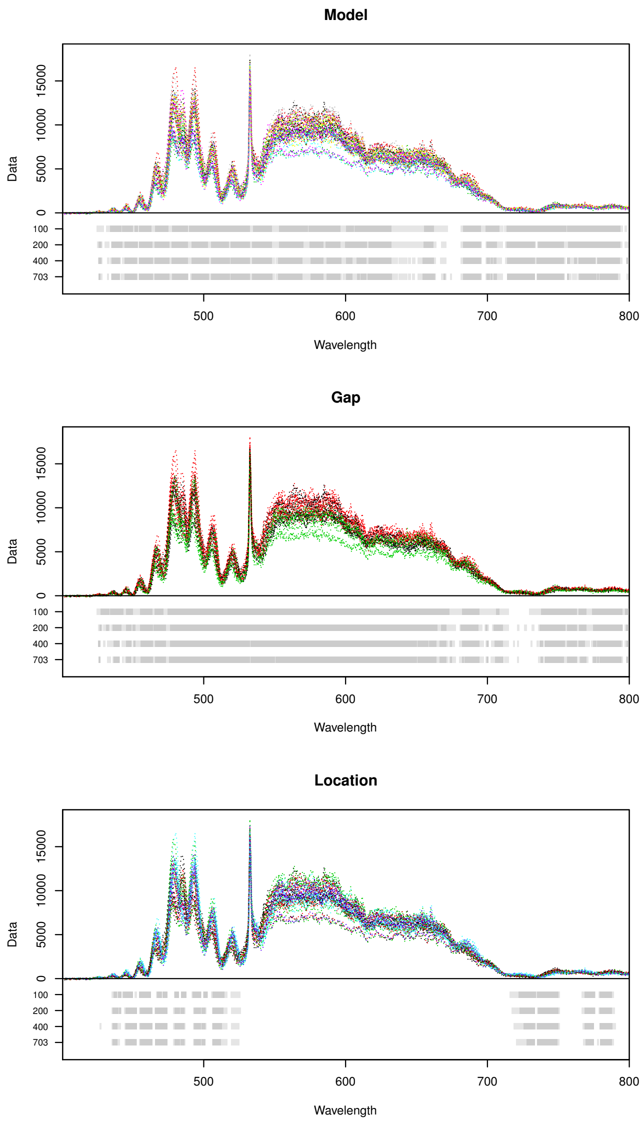

Figure 4: Emission data colored according to gap and location levels (top), gap levels (middle) and location levels (bottom). The gray bands below each graphic represent the results of tests on $\tilde{H}_{0,Model}$, $H_{0,Gap}$ and $H_{0,Loc}$, resp., varying the number of knots, at 5% and 1% levels (light and dark grey, resp.)

Colombo, D. and Previtali, B. (2009), Fiber laser welding of titanium alloys and its monitoring through the optical combiner, *in* 'ICALEO 2009 Congress proceedings', pp. 620–629.

Colombo, D. and Previtali, B. (2010), Through optical combiner monitoring of fiber laser processes, *Int. J. Mater. Form.* **3**(1), 1123–1126.

Colosimo, B. M. and Pacella, M. (2007), On the use of principal component analysis to identify systematic patterns in roundness profiles, *Quality and reliability engineering international* **23**(6), 707–725.

Colosimo, B. M. and Pacella, M. (2010), A comparison study of control charts for statistical monitoring of functional data, *International Journal of Production Research* **48**(6), 1575–1601.

Cuesta-Albertos, J. A. and Febrero-Bande, M. (2010), A simple multiway anova for functional data, *TEST* **19**(3), 537–557.

Cuevas, A., Febrero, M. and Fraiman, R. (2004), An anova test for functional data, *Comput. Statist. Data Anal.* **47**(1), 111–122.

Daimler Chrysler AG (2005), 'Patent no. DE10241593'.

Del Castillo, E. and Colosimo, B. (2011), Statistical shape analysis of experiments for manufacturing processes, *Technometrics* **53**(1), 1–15.

Del Castillo, E., Colosimo, B. M. and Alshraideh, H. (2012), Bayesian modeling and optimization of functional responses affected by noise factors, *Journal of Quality Technology* **44**(2), 117–135.

Drignei, D. (2010), Functional anova in computer models with time series output, *Technometrics* **52**(4), 430–437.

Freedman, D. and Lane, D. (1983), A nonstochastic interpretation of reported significance levels, *J. Bus. Econ. Stat.* **1**(4), 292–298.

Grasso, M., Colosimo, B. M. and Pacella, M. (2014), Profile monitoring via sensor fusion: the use of pca methods for multi-channel data, *International Journal of Production Research* **52**(20), 6110–6135.

Gu, H. (2010), Laser lap welding of zinc coated steel sheet with laser-dimple technology, *J. Laser Appl.* **22**(3), 87–91.

Hall, P., Poskitt, D. S. and Presnell, B. (2001), A functional data analytic approach to signal discrimination, *Technometrics* **43**(1), 1–9.

He, Z., Zhou, P., Zhang, M. and Goh, T. N. (2014), A review of analysis of dynamic response in design of experiments, *Quality and Reliability Engineering International* .

Nair, V., Taam, W. and Ye, K. (2002), Analysis of functional responses from robust design studies, *Journal of Quality Technology* **34**(4), 355–370.

Noorossana, R., Saghaei, A. and Amiri, A. (2012), *Statistical analysis of profile monitoring*, John Wiley & Sons, New York.

Pini, A. and Vantini, S. (2015), Interval-wise testing for functional data, Technical Report 30/2015, MOX, Politecnico di Milano.

Ramsay, J. O. and Silverman, B. W. (2005), *Functional data analysis*, Springer, New York.

Schwoerer, T. (2008), Robot-guided remote laser scanner welding for highly-productive welding applications, *in* 'ICALEO 2008 Congress proceedings', pp. 392–398.

Steen, W. M., Mazumder, J. and Watkins, K. G. (2003), *Laser material processing*, Springer.

Yu, G., Zou, C. and Wang, Z. (2012), Outlier detection in functional observations with applications to profile monitoring, *Technometrics* **54**(3), 308–318.

MOX Technical Reports, last issues

Dipartimento di Matematica
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

29/2015 Antonietti, P.F.; Cangiani, A.; Collis, J.; Dong, Z.; Georgoulis, E.H.; Giani, S.; Houston, P.
Review of Discontinuous Galerkin Finite Element Methods for Partial Differential Equations on Complicated Domains

30/2015 Pini, A.; Vantini, S.
Interval-wise testing for functional data

28/2015 Taffetani, M.; Ciarletta, P.
Beading instability in soft cylindrical gels with capillary energy: weakly non-linear analysis and numerical simulations

27/2015 Marron, J.S.; Ramsay, J.O.; Sangalli, L.M.; Srivastava, A.
Functional Data Analysis of Amplitude and Phase Variation

25/2015 Del Pra, M.; Fumagalli, A.; Scotti, A.
Well posedness of fully coupled fracture/bulk Darcy flow with XFEM

26/2015 Tagliabue, A.; Dede', L.; Quarteroni, A.
Nitsche's Method for Parabolic Partial Differential Equations with Mixed Time Varying Boundary Conditions

24/2015 Bonaventura, L:
Local Exponential Methods: a domain decomposition approach to exponential time integration of PDE.

23/2015 Paolucci, R.; Mazzieri, I.; Smerzini, C.
Anatomy of strong ground motion: near-source records and 3D physics-based numerical simulations of the Mw 6.0 May 29 2012 Po Plain earthquake, Italy

22/2015 Bonaventura, L.; Ferretti, R.
Flux form Semi-Lagrangian methods for parabolic problems

21/2015 Arena, M.; Azzone, G.; Secchi, P.; Vantini, S.
Reputational risk and corporate finance: A multi-event model