Available online at www.sciencedirect.com

ScienceDirect

Procedia CIRP 99 (2021) 307–312

www.elsevier.com/locate/procedia

14th CIRP Conference on Intelligent Computation in Manufacturing Engineering, CIRP ICME '20

Binary Gaussian Process classification of quality in the production of aluminum alloys foams with regular open cells

Alfredo Anglani[a], Massimo Pacella[a,*]

[a]*Department of "Ingegneria dell'Innovazione", University of Salento, Piazza Tancredi 7, 73100 Lecce – ITALY*

* Corresponding author. Tel.: +39 0832 297812; fax: +39 0832 297825. *E-mail address:* massimo.pacella@unisalento.it

Abstract

Aluminum alloys foams with homogeneous and regular open cells have been frequently proposed and used as support structures for catalytic applications. In this kind of application, the quality of produced metal foam assumes primary importance. This paper presents an application of a classifier algorithm to predict quality in the manufacturing process of aluminum alloy foams with homogeneous and regular open cells. A data analysis methodology of experimental data, which is based on Binary Gaussian Process Classification, is presented. The proposed method is a Bayesian classification method, which gets away from any assumptions about the relationship between process inputs (the geometric design parameters of the regular unit cells) and process output (probability to obtain defective foam). We demonstrate that the proposed methodology can provide an effective tool to derive a model for the prediction of quality. An investment casting process, via 3D printing of wax patterns, is considered throughout the paper. Despite this specific case study, the methodology can be exploited in different processes in which the assumptions of traditional statistical approaches could not be easily verified, e.g., additive manufacturing.

Peer-review under responsibility of the scientific committee of the 14th CIRP Conference on Intelligent Computation in Manufacturing Engineering, 15-17 July 2020.

Keywords: Aluminum alloys foams; Product quality classification; Gaussian process classification; Bayesian classification

1. Introduction

Predicting the quality of products can be considered a meaningful insight into the optimal design of an industrial manufacturing process. Quality prediction can be based on an experimental determination of the manufacturing process.

In [1], we discussed an experimental determination of an investment casting process for producing aluminum alloy foams with homogeneous and regular open-cells, which are employed as catalytic supports. In general, catalytic open-cell metal foams have to be characterized by a large void fraction or porosity, have to offer low resistance to thermal transport [2], to mass transfer [3], and needs to be easily coated with a catalytic layer on their surface via dip coating or other loading methods [4]. In this kind of application, the geometrical characterization of produced metal foam assumes primary importance. The shape of open cells influences both product performance and process performance, particularly processing time, which strongly depends on the designed shape [5].

Examples of open-cell metal foams, obtained by an investment casting process, are depicted in Fig. 1. Three cylindrical metal foams with prismatic unit cells are reported (two samples with runners, by which the liquid metal entered the mold cavity). Runners and the sprue gating, are then cut from the casting. After minor machining post-processing, the casting (identical to the original wax pattern) is ready for use.

In [1], experimental results on an aluminum casting alloy showed that the main limitations of investment casting concern the minimum wall thickness that can be cast to obtain a safe part. If this limit is exceeded, an incomplete filling of the mold will occur, as shown by the defective sample depicted in Fig. 2 (incomplete filling).

2212-8271
Peer-review under responsibility of the scientific committee of the 14th CIRP Conference on Intelligent Computation in Manufacturing Engineering, 15-17 July 2020.
10.1016/j.procir.2021.03.046

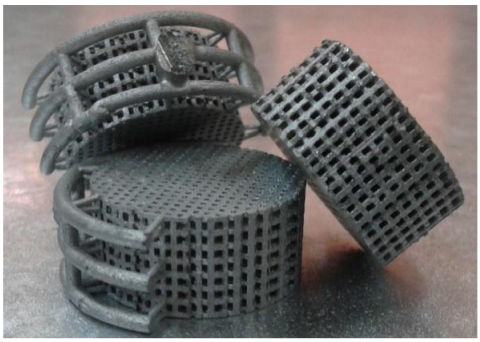

Fig. 1. Examples of cylindrical metal foams obtained by investment casting

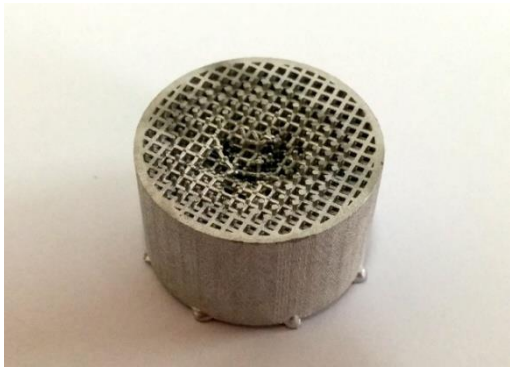

Fig. 2. Example of a defective foam sample obtained by investment casting

Another constraint concerns the maximum achievable total height of the foam sample, due to limitations of the casting furnace, assuming the use of a single sprue and several patterns attached. Hence, in [1] a statistical model for predicting quality of the investment casting process was derived based on geometric characteristics of the produced metal foam (wall thickness and total height). In particular, logistic regression and Response Surface Methodology were proposed for determining a statistical model of the manufacturing process.

In the present work, a supervised classification is proposed as an alternative for predicting quality in the process. In supervised classification, a set of training data points with their corresponding classes are available to learn the underlying structure of the model. Based on this information, the objective is to infer the classes of new data points.

Kernel machines have recently received much attention in the literature for classification problems. Examples of kernel classification algorithms include Support Vector Machine and Gaussian Processes (GPs). Herein, we consider a GP, which is a Bayesian kernel nonparametric probabilistic classification model originating from Gaussian process regression. Since the type of classification targeted is of 'defective/not-defective', we focus on binary GP classification (classes labeled as $-1/1$ respectively).

A latent function, distributed as a GP [6], is considered as the basis for binary GP classification. This latent function has a value, at a certain data point, which is related to the probability that this data point belongs to one class. The GP prior has some hyper-parameters that can be estimated by maximizing the log marginal likelihood [6]. When the hyper-parameters are estimated, the posterior distribution over the latent function evaluated at the training data points can be computed. This, in turn, allows us to predict the classes for new data points.

The remainder of this paper is organized as follows. In section 2, the investment casting process is presented. Section 3 describes the experimental procedure implemented. In section 4, the binary GP classification model is presented, while in section 5 the actual results are discussed. Finally, conclusions are provided in section 6.

2. The investment casting process

The term metal foam applies to porous structures whose solid matrix has a large fraction of interconnected, homogenous, and regular open cells. Typical applications of this material are catalytic supports, heat exchangers, filter elements, acoustic absorbers, crash absorbers, etc., because they have special properties, such as the permeability of the open cell structure, high porosity, and high ratio of surface area to volume. When the porosity cannot be subdivided into well-defined cells, the material is usually referred to as a metal sponge.

Investment casting has been shown as an effective technology for the production of digitally designed regular open-cell foams, as an inexpensive alternative to direct metal additive manufacturing. In the investment casting (a.k.a. lost-wax casting or micro-casting), the material is poured into a ceramic cavity designed to create an exact duplicate of the desired part using soluble preform (a prototype, a.k.a. wax pattern, usually of a polymer material, and obtained by a 3D printing process).

The ceramic cavity is obtained by investing (surrounding) the wax patterns of refractory material. The fragile wax patterns must withstand forces encountered during the mold making. After embedding, the mold is dried and then baked to harden the embedded material and to decompose (and evaporate) the polymer foam, leaving behind a negative image of the foam. Subsequently, the mold is filled with molten metal in a hot state under moderate pressure.

Aluminum alloys are usually used in investment casting mainly due to their high fluidity. After solidification and cooling, the mold materials are removed (e.g., by water under high pressure), leaving the metal foam that is an exact image of the original polymer prototype.

The main limitation of the investment casting technology is the spatial resolution currently offered. The limitation on the minimum wall thickness adds an obvious constraint also on the minimum cell size for a given level of porosity. This constraint might become cumbersome if small cell sizes are required. The experimental study in [1] showed that nearly any macroscopic geometrical property relative to catalytic applications (surface-to-volume ratio, hydraulic porosity, etc.) could be obtained by investment casting process.

3. Experimental determination of the process

The design process of a cylindrical foamed structure with a large fraction of interconnected, homogenous, and regular

open cells is guided by quality specifications such as density and material volume/wettable surface ratio. Computer-aided design (CAD) models of the actual metal foam can be used to yield a precise estimate of the true geometrical parameters. An example is given in Fig. 3, where three geometrical parameters are characterizing the cylindrical foamed structure [1]: (i) the wall thickness (a.k.a. shell size) (*S*), (ii) the height of the cylinder (*H*) and (iii) the open cell thickness (*tk*).

An experimental determination of the investment casting process consists of performing a set of metal foam casting tests, where the geometric characteristics of the foam (*S*, *H*, and *tk*) are changed. After each test, the observed quality of the metal foam is labeled as 'defective' or as 'not-defective' point in the set of process parameter combinations. In [1], starting from a set of experimental tests, a statistical analysis of collected tests was discussed. Statistical approaches are required to account for the randomness of the failure phenomenon. In particular, logistic regression, which allows estimating the relationship between some predictor variables (the foam geometric parameters *S*, *H*, and *tk*) and a dichotomous outcome (such as defective/not-defective), was used. In practice, logistic regression allows us to directly estimate the probability of defective (or 'failure process'), by considering both the defective and not-defective experimental points in the analysis.

When the response variable is dichotomous or ordinal, common linear models of Designs of Experiments, such as the standard second-order RSM designs, including the central composite design (CCD) and its variations (the rotatable CCD and the spherical CCD), the Box-Behnken design, etc., are inappropriate, and different approaches should be used to construct experimental designs. In [1], to handing a non-normal response of as in the logistic regression, a generalized linear model was fitted to the response data. In particular, a 15-run D-optimal design (no replication) for a GLM with a Bayesian approach was implemented. Table 1 reports the experimental design with the coded value for both *S* and *H* between -1 and 1.

The column *y* (the label -1/+1) is related to the experimental observation on the cylindrical metal foam obtained by design parameters *S* and *H*. The value $y=1$ represents a non-defective sample, while $y=-1$ a defective one.

The design was repeated for two levels of the open cell thickness (*tk*). The difference observed in the experimental results was related to the case $S = 0.35$mm $H = 30$mm and the foaming process failed for $tk = 0.4$mm, not for $tk = 0.5$mm.

Table 1 (first column) also reports the experimental order based on the *maxmin* distance criterion, in which space-filling of the experimental region is obtained by sequentially selecting the points in the region so that they are as far apart as possible from the points previously experimented in the sequence.

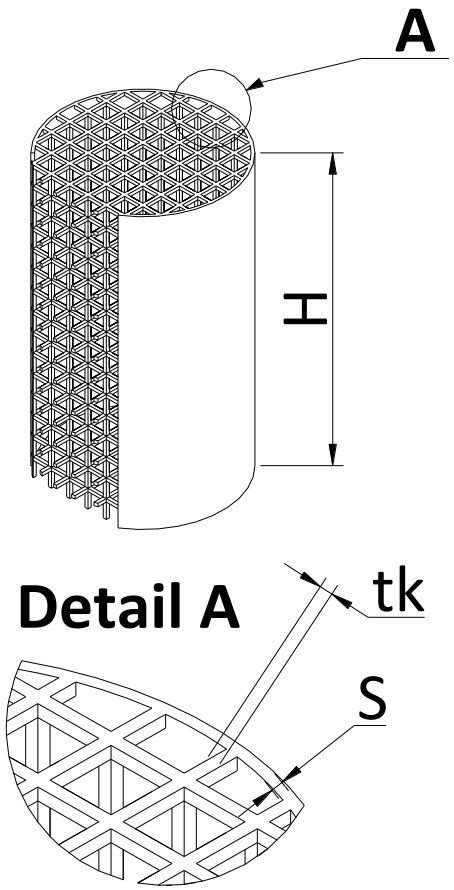

Fig. 3. A cylindrical foamed structure designed with CAD

The use of a Bayesian D-optimal design ensures good one-dimensional projections, whereas the *maximin* distance criterion ensures good space-filling in the full-dimensional space.

Table 1. 15-run Bayesian D-optimal design for *tk=0.4mm* [*tk=0.5mm*]

Order	*S (mm)*	*S coded*	*H (mm)*	*H coded*	*y*
10	1.00	1.000	45	0.309	1
3	0.25	-1.000	55	0.747	-1
12	0.45	-0.492	60	1.000	1
7	0.50	-0.290	25	-0.516	1
1	1.00	1.000	55	0.890	1
4	0.80	0.428	15	-0.965	1
8	0.30	-0.874	35	-0.008	-1
5	0.60	-0.127	45	0.282	1
9	0.80	0.373	35	-0.074	1
6	1.00	1.000	30	-0.365	1
2	0.25	-1.000	15	-0.990	-1
11	0.80	0.439	60	1.000	1
14	0.45	-0.471	45	0.238	1
15	0.35	-0.723	45	0.288	-1
13	0.35	-0.687	30	-0.426	-1[1]

4. Gaussian process for binary classification

Gaussian processes (GPs) for regression assume that the target function has a Gaussian process prior. Usually, the mean of this Gaussian is assumed to be a constant value, while the covariance between the targets at two different points is a decreasing function of their distance. This decreasing function is controlled by a small set of hyper-parameters that capture the interpretable properties of the function, such as the length scale of autocorrelation, the overall scale of the function, and the amount of noise. The posterior distributions of these hyper-parameters given the data can be inferred in a Bayesian way by maximizing the marginal likelihood.

While the target values in GP regression are continuous real values, in GP classification they are discrete class labels (in binary classification, two-class labels). Since it is not appropriate to assume that the target function with discrete outputs has a Gaussian process prior, it is assumed that there is some latent function whose value at a certain input location is monotonically related to the probability of belonging to a certain class at that location. As a consequence, the latent function, rather than the target function, has a Gaussian process prior. Starting from a prior on this latent function, the data points are used to infer both the posterior distribution over the latent function and the values of hyperparameters, which determine various aspects of the function.

Consider binary observations, $Y=\{y_i\}, y_i \in \{-1,1\}$, appointed to $X=\{\mathbf{x}_i\}, \mathbf{x}_i \in \Re^D$, $\mathcal{D}=\{(\mathbf{x}_i, y_i)\}, i=1,\cdots,N$. The observations are considered to be drawn from a Bernoulli distribution with a success probability $p(y_i=1|\mathbf{x}_i)$. Given this data set, the classification problem is to output the correct class label for a new data point.

To represent our uncertainty over class labels, one may want a method that outputs probabilities over the different labels for each new data point. We assume that the probability over class labels as a function of $\mathbf{x}$ depends on the value of the latent real-valued function $f(\mathbf{x}): \Re^D \to \Re$, which is mapped to a unit interval by a sigmoid transformation g as follows (a.k.a. logistic model, which we used in this study).

$$g(z)=\left(1+\exp(-z)\right)^{-1}. \tag{1}$$

The transformation used is a function $p(y_i=1|\mathbf{x}_i)=g(f(\mathbf{x}_i))$, where g is also called the link function. Hence, given the latent function $f(\mathbf{x})$, for binary classification, the probability of class label is independent of all other quantities. We put a Gaussian process prior on the latent function, which implies that any finite subset of latent variables has a multivariate Gaussian distribution [5]. Namely, at the observed inputs the latent variables, $\mathbf{f}=[f(\mathbf{x}_i)]$, have a Gaussian prior distribution $p(\mathbf{f}|X)=N(\mathbf{f}|\boldsymbol{\mu},\mathbf{K})$, where $\mathbf{K}$ is the covariance matrix and $\boldsymbol{\mu}$ the mean function. While the prior on mean is set to a constant value, the covariance matrix is constructed by a covariance function $k(\mathbf{x}_i,\mathbf{x}_j|\Theta)$, which represents the prior assumptions of the smoothness of the latent function parameterized by Θ that we can learn from the data. The covariance function determines the covariance between two latent variables. i.e., how the response at one input is affected by the responses at another input.

The covariance function can be defined by various kernel functions. Different kernel functions lead to a different degree of smoothness and different structures. A widely used covariance function is the stationary squared exponential:

$$k(r|\Theta=\{\sigma_{se}^2, l_1,\cdots,l_D\})=\sigma_{se}^2\exp(-r^2), \tag{2}$$

$$r=\sqrt{\sum_{d=1}^{D}\left(x_{i,d}-x_{j,d}\right)^2/l_d^2}, \tag{3}$$

where r is the distance between two input vectors of index i and j, and σ_{se}^2 the magnitude parameter. The length-scale, l_d^2, governs how fast the correlation decreases among input dimension d. The process associated with squared exponential is indefinitely mean square differentiable, which is a strong assumption on the smoothness of $f(\mathbf{x})$. In our study, we used the covariance function in equation (2), as it is probably the most widely used in the machine learning literature [5].

Given the latent function, the class labels are independent Bernoulli variables and, we can write the conditional posterior of the latent function as

$$p(\mathbf{f}|\mathcal{D},\Theta)=\frac{p(\mathbf{f}|X,\Theta)}{p(\mathcal{D},\Theta)}\prod_{i=1}^{N}p(y_i|f(\mathbf{x}_i),\Theta). \tag{4}$$

Here, $p(\mathcal{D},\Theta)=\int p(y|F)\,p(F|X,\Theta)\,dF$ is the marginal likelihood of the hyper-parameters, and $p(F|X,\Theta)$ represents the Gaussian prior distribution of f, which can be written as follows.

$$p(\mathbf{f}|X,\Theta)=\frac{1}{(2\pi)^{\frac{n}{2}}|\mathbf{K}|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}(\mathbf{f}-\boldsymbol{\mu})^T\mathbf{K}^{-1}(\mathbf{f}-\boldsymbol{\mu})\right). \tag{5}$$

Finally, the class probability at a test point would be obtained by integrating over the hyper-parameters weighted by their posterior probability, as in the following equation.

$$p(\tilde{y}|\tilde{\mathbf{x}},\mathcal{D})=\int p(\tilde{y}|\tilde{\mathbf{x}},\mathcal{D},\Theta)\,p(\Theta|\mathcal{D})\,d\Theta. \tag{6}$$

This integral is costly and there are usually many fewer hyper-parameters than data points. Therefore, in this paper, rather than integrating over the hyper-parameters, we fit them by maximizing the marginal likelihood as $\hat{\Theta}=\arg\max_{\Theta} p(\mathcal{D}|\Theta)$ and predict using these best fit hyper-parameters: $p(\tilde{y}|\tilde{\mathbf{x}},\mathcal{D},\hat{\Theta})$.

The marginal likelihood $p(\mathcal{D},\Theta)$ and Gaussian prior distribution $p(\mathbf{f}|\mathcal{D},\Theta)$ in equation (4) are both analytically intractable due to the nonlinearities in (1). Therefore, approximation techniques need to be used to get the posterior. One widely used technique is the expectation propagation (EP) algorithm. The implementation of the EP used in our study was done via the Gaussian Process for Machine Learning (GPML) TOOLBOX in MATLAB [7]. A comparison of different methods for approximate inference in GP classification is provided in [8], it is showed that the EP

algorithm provides high accuracy and speed when compared to alternative methods.

5. Classification experiments

The GPML TOOLBOX was used for binary GP classification of the investment casting process based on 2 input factors: (i) x_1 the wall thickness foam (*S*) and (ii) x_2 the height of the cylinder (*H*). Value of *S* ranging between $0.25mm$ and $1.00mm$ was and of *H* ranging between $15mm$ and $60mm$ were considered. The design was replicated for two different values of the open cell thickness: $tk=0.4mm$ and $tk=0.5mm$.

A 15-run Bayesian D-optimal design, with no replication, was implemented (Table 1). Table 1 also reports the coded value for both *S* and *H* ranging between -1 and 1. Table 1 also reports the experimental order, based on the *maxmin* distance criterion, in which space-filling of the experimental region is obtained by sequentially selecting the points in the region so that they are as far apart as possible from the points previously experimented in the sequence. While the *maximin* distance criterion ensures good space-filling in the full-dimensional space, the Bayesian D-optimal design ensures good one-dimensional projections.

In this study, we implemented a sequential sampling strategy [9]. In particular, four scenarios with a different number of tests (i.e. 6, 9, 12, and 15) were considered. The selection of tests in each scenario is based on the sequential order obtained by the *maxmin* distance criterion. Hence, the first 6 tests are included in the first scenario, the first 9 in the second, the first 12 in the third, and all of the 15 tests in the last scenario. In practice, instead of assuming that the experimental test data is already available, this work aims to sample experimental points sequentially to explore increasingly the design space and to construct a surrogate model in each scenario.

Fig. 4 shows the results of the GP classification based on the first 6 experimental tests. The input points are represented by 6 cross-marks. The resulting classification is displayed in form of a contour plot over the design space spanned by the factors *S coded* (horizontal axis) and *H coded* (vertical axis). For each of the two factors, the design space has been expanded to cover values greater than 1 and lesser than -1, i.e. ranging between -1.5 and 1.5. Fig. 5 and Fig. 6 show the results of the GP classification based on the first 9 and 12 experimental tests, respectively. A color scale is used to represent the predicted class probability $p\left(\tilde{y}=1\,|\,\tilde{x},\mathcal{D},\hat{\Theta}\right)$ at a test point by maximizing the marginal likelihood as $\hat{\Theta}=\arg\max_{\Theta} p\left(\mathcal{D}\,|\,\Theta\right)$. Dark colors indicate a low probability to obtain a non-defective product, while brighter ones are used for a higher probability of obtaining a non-defective product.

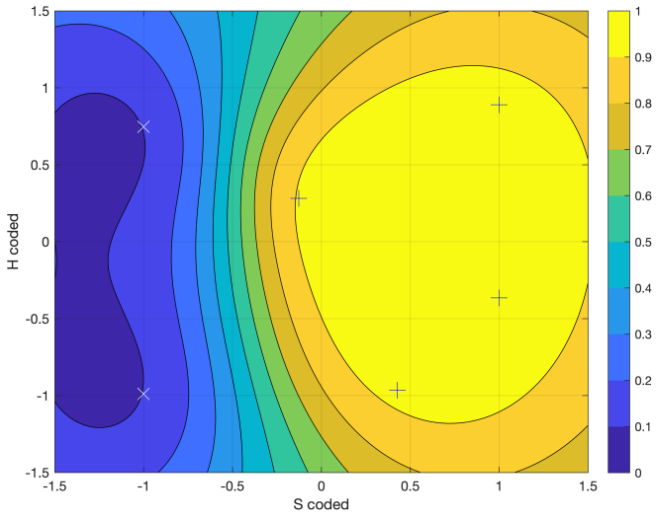

Fig. 4. GP classification based on the first 6 experimental tests

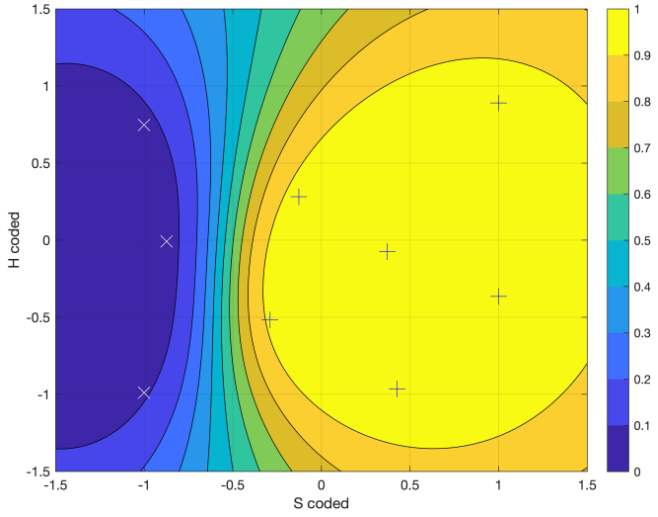

Fig. 5. GP classification based on the first 9 experimental tests

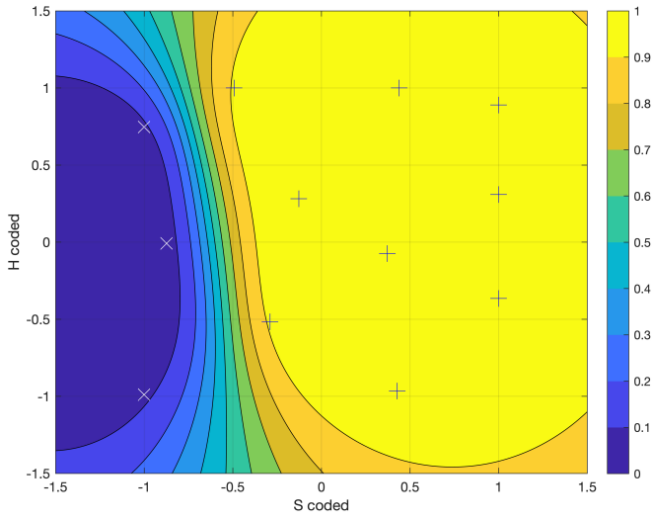

Fig. 6. GP classification based on the first 12 experimental tests

Fig. 7 and Fig. 8 show similar results when all of the tests in the 15-run Bayesian D-optimal design are considered. Fig. 7 refers to the case of open-cell thickness equal to $tk=0.4mm$ while Fig. 8 refers to $tk=0.5mm$.

It can be observed that the GP classifier can model a response variable in a sequential sampling strategy, and the classification is robust to the number of tests. This result assumes importance in cases of experiments that are too costly or time-consuming to conduct [9]. Furthermore, the GP classifier allows extending the prediction in points not included in the boundaries of the input design space. From Fig. 7 $(tk=0.4mm)$ it appears that values of the coded variable *S* greater than -0.5 (i.e., actual size of *S* greater than $0.44mm$) allow obtaining a safe result in the process [1]. In the case of $tk=0.5mm$ (Fig. 8), the GP classifier can model more complex interaction relationships between the two input factors that logistic regression in [1] is not able to model.

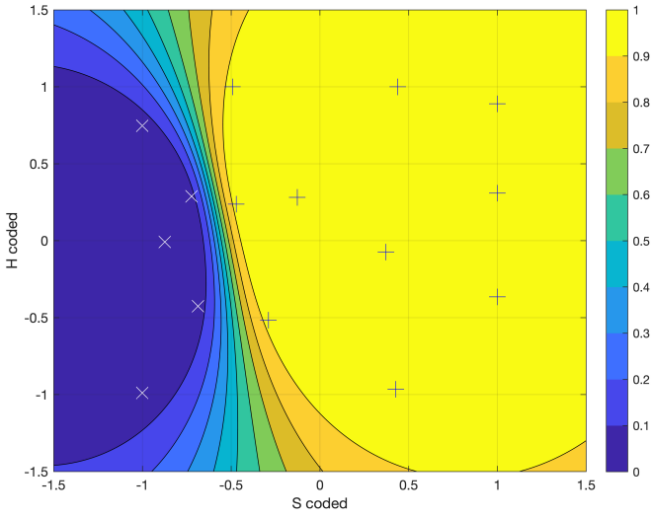

Fig. 7. GP classification based on all of the 15 tests. *tk=0.4mm.*

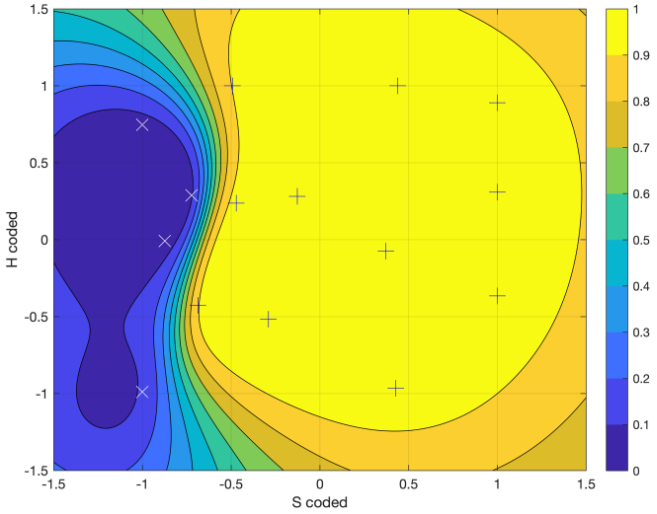

Fig. 8. GP classification based on all of the 15 tests. *tk=0.5mm.*

6. Conclusions

In this paper, a methodology based on a GP for classifying and predicting the quality of a manufacturing process has been presented. The proposed method is a Bayesian classification method, which gets away from any assumptions about the structural relationship between inputs and output. The method can model a response variable in a sequential sampling strategy. Furthermore, given new data, we can get a class probability rather than a hard decision.

A case study related to investment casting for the production of metal foams has been considered. In particular, the geometric characteristics of the foam (*S*, *H*, and *tk*) are used as the input factors of the manufacturing process, which may influence on the successful production. Despite this specific case study, the methodology can be exploited in different processes in which the assumptions of traditional statistical approaches for experimental data modeling could not be easily verified.

One advantage that we can observe on using GP classifiers for quality prediction is that they are fully non-parametric statistical models. Hence, we can improve the performance of GP classifiers from many different directions. Firstly, we can use the evidence for model selection and kernel hyperparameter optimization. Secondly, prior information can be used to inform the learning of the hyperparameters (for example, if some input features are thought to be more relevant). Finally, we can improve approximate inference and optimization techniques to gain in both accuracy and speed.

Acknowledgments

This work has been partially funded by the Ministry of Education, University and Research of Italy (MIUR).

References

[1] Anglani A, Pacella M, Logistic Regression and Response Surface Design for Statistical Modeling of Investment Casting Process in Metal Foam Production. Procedia CIRP 2018;67:504-509.
[2] Zhao CY, Review on thermal transport in high porosity cellular metal foams with open cells. Int J of Heat and Mass Transfer 2012;55: 3618-32.
[3] Ji X, Xu J, Experimental study on the two-phase pressure drop in copper foams. Heat and Mass Transfer 2012;48:153-64.
[4] Cristiani C et al., Activation of metallic open-cell foams via washcoat deposition of Ni/MgAl2O4 catalysts for steam reforming reaction. Catalysis Today 2012;197:256-64.
[5] Primo T, Calabrese M, Del Prete A, Anglani A, Additive manufacturing integration with topology optimization methodology for innovative product design. Int J of Adv Manufacturing Technology 2017;93:467-79.
[6] Rasmussen CE, Williams CKI, Gaussian Processes for Machine Learning. MIT Press, Cambridge, MA; 2006.
[7] Rasmussen CE, Nickisch H, Gaussian processes for machine learning (GPML) toolbox. J Machine Learning Research 2010; 11:3011-15.
[8] Nickisch H, Rasmussen CE, Approximations for binary Gaussian process classification. J Machine Learning Research 2008;9:2035-78.
[9] Gahrooei MR, Paynabar K, Pacella M, Colosimo BM, An adaptive fused sampling approach of high-accuracy data in the presence of low-accuracy data. IISE Transactions 2019;51:1251-64.