A computer-aided methodology for the optimization of electrostatic separation processes in recycling

Matteo Borrotti[1],* Antonio Pievatolo[1], Ida Critelli[2],
Andrea Degiorgi[2], and Marcello Colledani[2]
[1] *Institute of Applied Mathematics and Information Technology, National Research Council of Italy, Milan, Italy and*
[2] *Department of Mechanical Engineering, Politecnico di Milano, Milan, Italy*

(Dated: July 8, 2015)

Abstract

The rapid growth of technological products has led to an increasing volume of waste electrical and electronic equipments (WEEE) which could represent a valuable source of critical raw materials. However, current mechanical separation processes for recycling are typically poorly operated, making it impossible to modify the process parameters as a function of the materials under treatment, thus resulting in untapped separation potentials. Corona Electrostatic Separation (CES) is one of the most popular processes for separating fine metal and nonmetal particles derived from WEEE. In order to optimize the process operating conditions (i.e. variables) for a given multi-material mixture under treatment, several technological and economical criteria should be jointly considered. This translates into a complex optimization problem that can be hardly solved by a purely experimental approach. As a result, practitioners tend to assign process parameters by few experiments based on a small material sample and to keep these parameters fixed during the process life-cycle. The use of *computer experiments* for parameter optimization is a mostly unexplored area in this field. In this work, a computer-aided approach is proposed to the problem of optimizing the operational parameters in CES processes. Three metamodels, developed starting from a multi-body simulation model of the process physics, are presented and compared by means of a numerical and simulation study. Our approach proves to be an effective framework to optimize the CES process performance. Furthermore, by comparing the predicted response surfaces of the metamodels, additional insight into the process behavior over the operating region is obtained.

Keywords: Design of computer experiments; Kriging; Artificial neural networks; Response surface; WEEE

*matteo.borrotti@mi.imati.cnr.it

I. INTRODUCTION

In recent years, the use of rules that govern experiments for technological improvement (i.e. Design of Experiments - DoE) [1] has received renewed momentum through the utilization of *Computer Experiments* [2]. These experiments are run on a computer code implementing a simulation model of a physical system of interest such as in the process of recovering waste printed circuit boards or in enzyme kinetics. In general a simulation model, or simulator, consists of a set of many linear or nonlinear, ordinary and/or differential simultaneous equations, whose solutions may not be amenable to analytical expression. Furthermore, runs of the simulator can be expensive and/or time-consuming. In this case, the use of a surrogate model (metamodel) is suggested. These metamodels are simpler models which represent a valid approximation of the original simulator and usually are statistical interpolators built from the simulated input-output data. Predictions at untried experimental points can then be made by metamodels and optimization techniques can be applied in order to find best settings. The selection of a metamodel to approximate the true model as accurately as possible is a crucial problem. Generally, polynomial surrogates are widely used to model computer experiments [3]. However, the behavior of the data often cannot be explained by these models. An alternative method is known as the neural network technique which allows the data to be fitted in a less constrained form [4, 5]. Another popular technique is kriging, which has been widely used for the design and analysis of computer experiments [6, 7]. With the aim of a detailed comparison between metamodels, in this work we have analyzed the output of a specific experimental design with three different techniques (i.e. polynomial regression, kriging and neural network models).

This analysis is applied to the Corona Electrostatic Separation (CES), which is a process widely used in recycling for separating conductive from non-conductive fine particles obtained from the shredding of waste electrical and electronic equipments (WEEE). The main goal is to understand advantages and disadvantages of each model in support of the optimization of the process parameters, ultimately aiming at implementing an in-line process control framework. CES is defined as the selective sorting of charged or polarized bodies in an electric field [8–10] and it represents one of the most effective ways for recycling metals and nonmetals from WEEE with limited environmental impact. For example, this process has been proposed as a viable alternative to pyrometallurgical processes in the recovery of

metals from waste Printed Circuit Board (PCB).

WEEE is currently considered to be one of the fastest growing waste streams in the world, growing at 3-5 % per year [11]. Electronic waste can represent a very important source of key-metals for advanced technological products. For example, PCBs are called urban mineral resources, as 25% - 30% of their composition in weight is made of valuable metals such as copper, tin, nickel, gold and silver. However, in order to successfully apply CES processes to the separation of complex multi-material mixtures such as granular flows from finely shredded PCBs, a tight control of the process parameters needs to be implemented. In order to optimize the process parameters (i.e. variables) for a given multi-material mixture under treatment, several technological end economical criteria should be jointly considered. This translates into a complex optimization problem that can hardly be solved by a purely experimental approach. As a result, practitioners tend to assign process parameters by means of few experiments based on a small material sample and to keep these parameters fixed during the process life-cycle. This is mainly due to the lack of knowledge-based engineering models and tools to support the design and operation of separation processes able to capture, with an acceptable level of confidence, all the major phenomena affecting the quality of the output. In the literature attempts at implementing different control procedures have been developed in order to optimize specific objectives functions [12, 13] based on simplified process models and computer experiments [14, 15].

In this study, a reliable multi-body simulation model of roll-type Corona Electrostatic Separation (CES) [16] is used to build a set of metamodels in order to conduct computer-aided experiments for optimizing the process parameters. This solution will provide the ability to optimize the process parameters with respect to the specific material mixture under treatment, thus providing a degree of adaptability to these processes that is currently not available in the recycling industry.

The paper is structured as follows. The engineering problem is described in Section 2. Then, in Section 3, three metamodels for the response of physical experiments are introduced and discussed. Section 4 presents the computer-aided approach to generate sampling plans and its application is illustrated by the case study based on the CES simulator. The predictive abilities of the three metamodels are investigated and compared. A final discussion concludes the paper in Section 5.

II. ENGINEERING PROBLEM

Corona electrostatic separators are mainly used to separate conductors from insulators, like copper from plastics, in shredded multi-material mixtures. The process physics is briefly explained in the following. Particles are transported by a feeder on the rotating drum and they are charged as they pass through an electrostatic field. More specifically, the separator uses corona charging or ion discharging to establish a negative charge on particles when particles pass between two high-voltage electrodes. Particles receive a discharge of electricity, which gives the non-metals a high surface charge, causing them to be attracted to the rotor surface until they are brushed down into the left bin. Metal particles do not get charged, as the charge rapidly dissipates through the particles to the earthed rotor, so they fall into the right bin under the effect of the centrifugal and gravitational forces. Ideally, if the separation process were perfectly accurate, all the particles would be correctly classified. However, due to the presence of mixed non-liberated particles and to random particle impacts, a non-classified material flow is generated (middlings) consisting of particles that drop in the central bin. The quality of the separation process depends on two sets of process parameters, namely design parameters and controllable parameters. The design parameters include the drum diameter, the position and the shape of the electrodes. The controllable parameters include the drum rotational speed, the electrostatic potential, the feed rate, and the splitters position. Moreover, the distribution of the input material characteristics (shape, density, conductivity, degree of liberation) also affects the separation performance. Due to these complex particle flow dynamics, the particles trajectories are far from being deterministic, but a statistical distribution of the particles falling points at the bin level is usually observed.

Since the process physics is complex, modeling the trajectories of particles within CES is of practical engineering interest. In the literature, several attempts have been made to simulate particle trajectories in roll-type electrostatic separators [17, 18]. However, existing recycling process models only focus on single particle trajectories and fail to model two major causes for loss of efficiency in the separation process, i.e. (i) particle-particle and particle-equipment interactions and impacts, and (ii) the presence of un-liberated particles in the material flow. These limitations undermine the applicability of existing models in industrial settings. More realistic models are needed to better capture the real behavior of the process and to provide an accurate prediction of separation process performance. These motivations

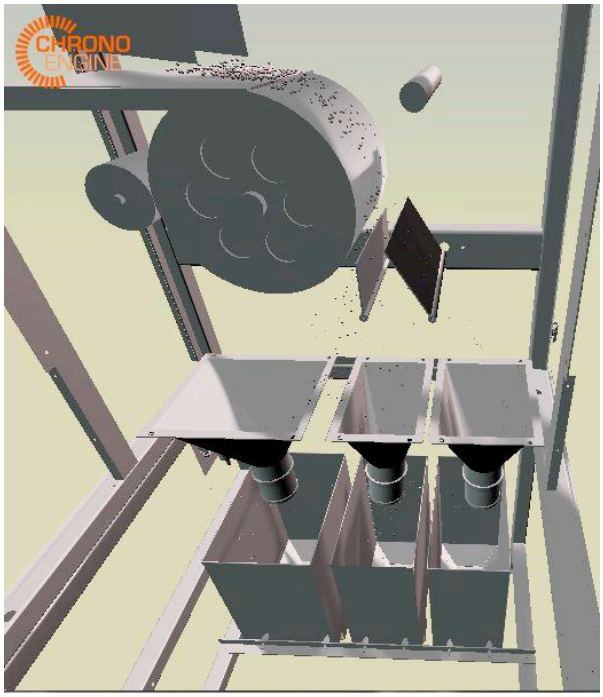


FIG. 1: Example of particle behaviour simulated by the CES simulation model.

has led to the development of a new multi-body simulation model [16, 19]. The simulation is based on the Chrono::Engine (`http://projectchrono.org`) C++ simulation libraries, which implement the Differential Variational Inequalities (DVI) approach. The non-smooth dynamics method used for the simulator, rooted in the recent theory of DVI, can handle up to millions of contacts between particles without requiring short integration time steps [20]. Within the simulation, each particle is randomly generated with its own characteristics by sampling from a particle flow "DNA", characterized by multiple attributes, such as shape, material, orientation, and liberation degree. External forces, including aerodynamic, electrical, centrifugal and gravitational forces are applied to these particles, determining their trajectories. These realistic simulations, that have been validated in real industrial settings, can support the designer to perform virtual experiments to test the behavior of the system under specific mixture conditions and machine parameters, instead of running real experiments. From an application point of view, this approach can reduce drastically the number of real experiments and make the process flexible and adaptable to changes within the material flow. An example of separation process computed by the CES simulator is reported in Fig. 1.

However, simulations are usually time-consuming and it is still unpractical to rely on them

for implementing in-line process control schema. To this purpose, the use of metamodeling is particularly appealing for enabling "fast" in-line applications.

In this work we consider a typical industrial CES machine (Hamos KWS electrostatic separator) with fixed design parameters, whose performance depends on the following controllable parameters:

1) Electrostatic potential, or voltage, which ranges between -35.000 to -25.000 Volt;

2) Drum speed, or simply speed, which ranges between 32 to 128 rpm;

3) Feed rate, which ranges between 0.0028 to 0.028 kilograms per seconds (kg/s).

Furthermore the position of two splitters, S_1 and S_2, separating the collecting boxes of conductive, middling and nonconductive products are considered. They have a fixed height and can move along the horizontal coordinate with a range of variation between 0 and 20 cm for both, divided into 60 possible positions. In the case under analysis the splitters are not considered as physical bodies, unlike the particles or the drum, and the interaction of the particles with these volumes is neglected. The positions of S_1 and S_2 are only considered in the post-processing phase: knowing the coordinates of the falling point of each particle after each simulation, the best position that maximizes an objective function is found, testing all feasible combinations of S_1 and S_2. The outcomes (i.e. system responses) of the process are: the recovery of conductive products in its collecting box ($R_{c,c}$), the grade of conductive products in its collecting box ($G_{c,c}$), the recovery of non-conductive products in its collecting box ($R_{nc,nc}$) and the grade of non-conductive products in its collecting boxes ($G_{nc,nc}$). For the case of conductive products, the recovery is defined as follows:

$$R_{c,c} = \frac{m_{c,c}}{m_{c,c} + m_{c,m} + m_{c,nc}}, \tag{1}$$

where $m_{c,c}$ is the mass of conductive products in its collecting box, $m_{c,m}$ is the mass of conductive products in the middling box and $m_{c,nc}$ is the mass of conductive products in the non-conductive box. Instead, the grade is:

$$G_{c,c} = \frac{m_{c,c}}{m_{c,c} + m_{nc,c}}, \tag{2}$$

where $m_{nc,c}$ is the mass of non-conductive products in the conductive box. Similarly, $R_{nc,nc}$ and $G_{nc,nc}$ are defined as the mass of non-conductive products in its collecting box divided

by the sum of the mass of non-conductive products in all the boxes and the mass of non-conductive products in its collecting box divided by the total mass in the non-conductive box, respectively.

These system responses are in a trade-off and different possible solutions can arise while searching for a process optimization. For instance an optimal solutions for recovery $R_{c,c}$ can negatively influence the performance of the separation process in terms of grade $G_{c,c}$. For this reason, we selected a utility function, which assesses the material recycling potential of a product. This approach is similar to Sherwood's characterization of the relationship between the price of a material and its concentration in its feed stream [11, 21], described in Gutowski et al. (2008) [22]. The utility function, u, has been defined as:

$$u(\phi_1, \phi_2) = \phi_1(R_{c,c}, G_{c,c}) + \phi_2(R_{nc,nc}, G_{nc,nc}), \tag{3}$$

where:

$$\phi_1(R_{c,c}, G_{c,c}) = (3.0649 \times (G_{c,c})^2 + 1.5747 \times G_{c,c} + 1.6673) \times R_{c,c} \tag{4}$$

and

$$\phi_2(R_{nc,nc}, G_{nc,nc}) = 0.1 \times (3.0649 \times (G_{nc,nc})^2 + 1.5747 \times G_{nc,nc} + 1.6673) \times R_{nc,nc}. \tag{5}$$

The constant values are chosen in order to fit copper economic values.

The utility function $u(\phi_1, \phi_2)$ is then used to calculate the final system response, y_i with $i = 1, \ldots, E$, where E is the total number of experimental points given by all controllable parameters combinations, and should be maximized in order to optimize the electrostatic separation process. $R_{c,c}$, $G_{c,c}$, $R_{nc,nc}$ and $G_{nc,nc}$ are computed using the simulators of the CES process [16], previously described.

III. METAMODELS

A. Polynomial regression model

Polynomial regression models [23] are often used to approximate the dependence relation between a set of variables and a response. This response represents the output of the system and is measured by an identified variable Y and the v input variables $(x_1, x_2, \ldots, x_v)$ describe the features of the problem. This relation can be written as follows:

$$Y = f\Big(x_1, x_2, \ldots, x_v\Big) + \epsilon; \tag{6}$$

where f may be a smooth function of $x_1, x_2, \ldots, x_v$ and ϵ represents a random noise in the observable response.

The simplest polynomial model to explore the space approximating the function (6) is the first-order polynomial model,

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_v x_v + \epsilon; \tag{7}$$

where v variables are assumed to affect the response in a linear way without interactions. The first-order polynomial model is likely to be appropriate when we are interested in approximating the true response surface over a relatively small region of the independent variable space in a location where there is little curvature in f. Often the curvature in the true experimental surface is strong enough that the first-order model is inadequate. A second-order polynomial model will likely be required in these situations. The second-order polynomial model is a flexible model to describe experimental data in which nonlinear terms are present. The nature of the experimental surface depends on the signs and magnitudes of the coefficients in the following model:

$$Y = \beta_0 + \sum_{i=1}^{v} \beta_i x_i + \sum_{i=1}^{v} \beta_i x_i^2 + \sum_{i<j} \beta_{ij} x_i x_j + \epsilon \tag{8}$$

The approach of least squares estimation for the β parameters is then generally used and the adequacy of the fitted model is evaluated with the typical linear model diagnostics, such as residual analysis, coefficient of determination and graphical methods.

B. Artificial Neural Network (ANN)

An Artificial Neural Network (ANN) [24] is an information processing paradigm that is inspired by the way biological nervous systems, such as the brain, process information. The structure of the information processing system is composed by a large number of highly interconnected processing elements (neurons) working in unison to solve specific problems. An ANN is configured through a learning process, which, as in biological systems, involves adjustments to the synaptic connections that exist between the neurons.

The usual type of ANN consists of three groups, or layers, of units: a layer of *input units* (i.e. variables) is connected to a layer of *hidden units*, which is connected to a layer

of *output units* (system responses). When all the units are connected to one another and information travels one way only, from input to output, the ANN has a single-layer feed-forward architecture as represented in Fig. 2. Each hidden layer can be composed by a different number of nodes.

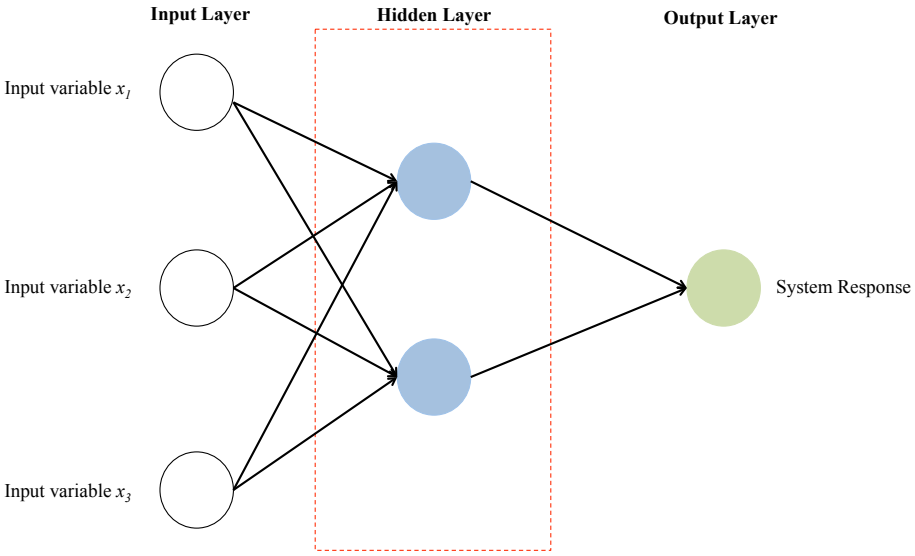


FIG. 2: Example of a single layer feed-forward neural network.

The influence of the inputs on the outputs is determined by the connection weights. The connection weights are real numbers associated with each connection between units and they determine whether it is possible for one unit to influence another. In order to train an ANN, the weights of each unit should be adjusted in such a way that the error between the expected output and the actual output is reduced. The most widely used method to determine the best weights is the back-propagation algorithm [25], which minimizes this error by means of a gradient descent method that selects the best combination of weights. The performance of the algorithm depends on a weight decay parameter, λ, which is a regularization term that tends to decrease the magnitude of the weights and avoids over fitting [24].

After the training, the output of an ANN is equal to:

$$Y = f\Big(\sum_{i=1}^{v} w_i \times x_i\Big); \tag{9}$$

where w_i is the connection weight of the input variable x_i, with $i = 1, \ldots, v$. The function f, which is called activation function or transfer function, is chosen before the training among

a range of functions [26]. Here we have employed the sigmoid function defined as follows:

$$f(x) = \frac{1}{1 - \mathrm{e}^{-x}} \tag{10}$$

where x is a real number.

C. Kriging Model

The kriging methodology was originally proposed by a South Africa geologist, D. G. Krige (1951) [27], for the analysis of geostatistical data and, nowadays, has become very popular in several applied contexts [28, 29].

Following Sacks et al. (1989) [30], the kriging approach consists in treating the output of a simulator as a realization of a stochastic process $Y(\mathbf{x})$ such that $Y(\mathbf{x}) = \mu(\mathbf{x}) + Z(\mathbf{x})$. The global trend is denoted by $\mu(\mathbf{x})$ and $Z(\mathbf{x})$ represents the departure of the system response $Y(\mathbf{x})$ from the trend. More precisely, $Z(\mathbf{x})$ is usually assumed to be a Gaussian stationary process with $E(Z(\mathbf{x})) = 0$, a constant variance σ_Z^2, and a non-negative correlation function between two experimental points $\mathbf{x}_1$ and $\mathbf{x}_2$:

$$Corr[Z(\mathbf{x}_1), Z(\mathbf{x}_2)] = R(\mathbf{x}_1, \mathbf{x}_2). \tag{11}$$

The correlation function should reflect the characteristics of the system response and one of special interest is the *power exponential family*, which is defined coordinatewise to form the correlation function in the v-dimensional space under the assumption of separability:

$$R(\mathbf{x}_1, \mathbf{x}_2|\psi) = \exp\left\{ -\sum_{i=1}^{v} \left| \frac{x_{1i} - x_{2i}}{\theta_i} \right|^{p_i} \right\}, \tag{12}$$

where, for $i = 1, \ldots, v$, θ_i is a scale parameter, p_i is the power parameter and ψ is the vector having $\psi_i = (\theta_i, p_i)$ as its i-th entry. A popular choice for the estimation of ψ is the maximum likelihood estimate (MLE).

Two different types of kriging metamodels should be considered depending on the functional form of the trend function [7]:

- Ordinary kriging: the trend is constant $\mu(\mathbf{x}) = \mu$ but unknown.

- Universal kriging: the trend component depends on $\mathbf{x}$ and is modeled in a regressive way

$$\mu(\mathbf{x}) = \sum_{i=1}^{v} f_i(\mathbf{x})\beta_i, \tag{13}$$

where $f_1(), \ldots, f_v()$ are known functions and $\beta = (\beta_1, \ldots, \beta_v)$ is the vector of unknown parameters.

In other words, in ordinary kriging, the trend is a simple unknown constant whereas in universal kriging the trend varies and the coefficients of the regression equation that describes this trend are unknown. In this work, we have employed the universal kriging modeling approach with the simplest possible trend, that is, a first-order polynomial, letting $f_i(\mathbf{x}) = x_i$.

IV. COMPUTER-AIDED EXPERIMENTAL APPROACH

Following the general framework for computer experiments we proposed a computer-aided experimental approach (see Fig. 3) which is based on the following main steps:

(i) Select the initial experimental batch of size n based on a sampling-based design [2];

(ii) Evaluate the design points with the simulator;

(iii) Estimate the metamodel on the simulated input-output data;

(iv) Select from the experimental region the best point which yields the highest predicted system response value;

(v) Evaluate the best design point with the simulator and use it as the best setting.

A. Sampling-based design: Latin Hypercube Sampling (LHS)

Computer Experiments differ from traditional physical experiments in that repeated observations at the same set of input variables yield identical system response [2], therefore, sampling techniques that spread design points are needed. Different sampling-based designs are described in the literature [31] and one of the most used is the Latin Hypercube Sampling (LHS) technique [32] or Latin Hypercube Design. Generally, to obtain an LHS design

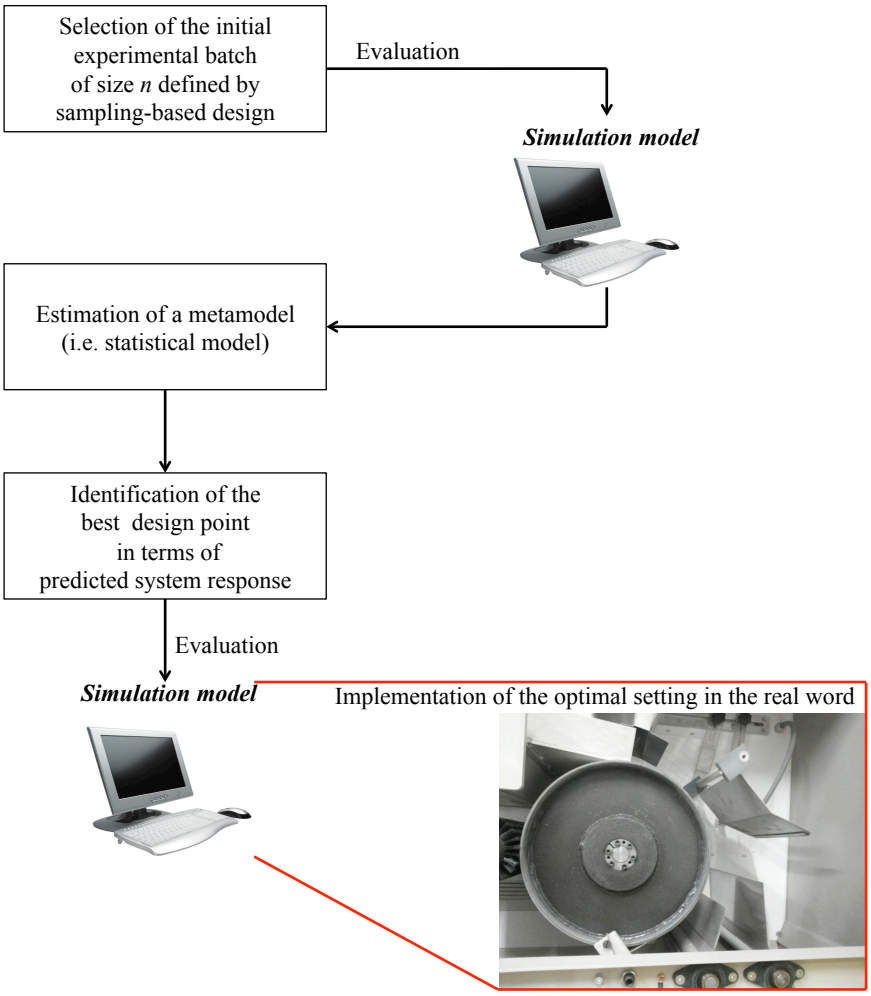


FIG. 3: Computer-aided experimental approach.

consisting of n points, each dimension of the experimental region is divided into n equally spaced intervals (e.g. $[0, (1/n)], \ldots, [(n-1)/n, 1]$), an interval is randomly sampled without replacement from each dimension and a value is uniformly sampled from the interval, obtaining one design point. This procedure is then repeated n times (see Fig. 4).

The uniform sampling of the co-ordinates of the design points meets the S-optimality criterion, which consists in maximizing the harmonic mean distance from each design point to all the other points in the design [33], in order to spread them as much as possible in the experimental region.

In this work, the Latin hypercube design is applied and a batch of size $n = 50$ is selected based on the S-optimal criterion. The batch is then tested on the CES simulator.

As a first exploratory analysis, a set of descriptive statistics have been computed in order

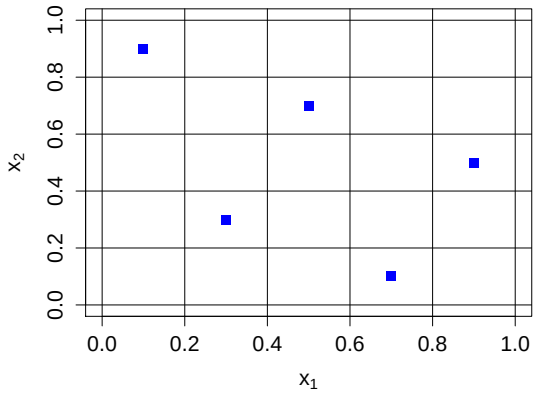


FIG. 4: A Latin hypercube design with two variables, x_1 and x_2.

to get some insights into the frequency distribution of the system response. The maximum value of the utility function reached with the Latin hypercube design is 6.928 and the minimum is 4.383. The mean is 5.866 and the median is equal to 5.904; the first quartile is 5.319 and the third quartile is 6.358.

These values indicate a slightly left-skewed distribution and this particular shape can be seen in Fig. 5(a). In Fig. 5(b) a box-plot is shown which summarizes the distribution of the system response. The upper side of the box is the third quartile, instead the lower side of the box is the first quartile. The line inside the box represents the median. The upper and lower whisker are, respectively, the maximum and minimum values.

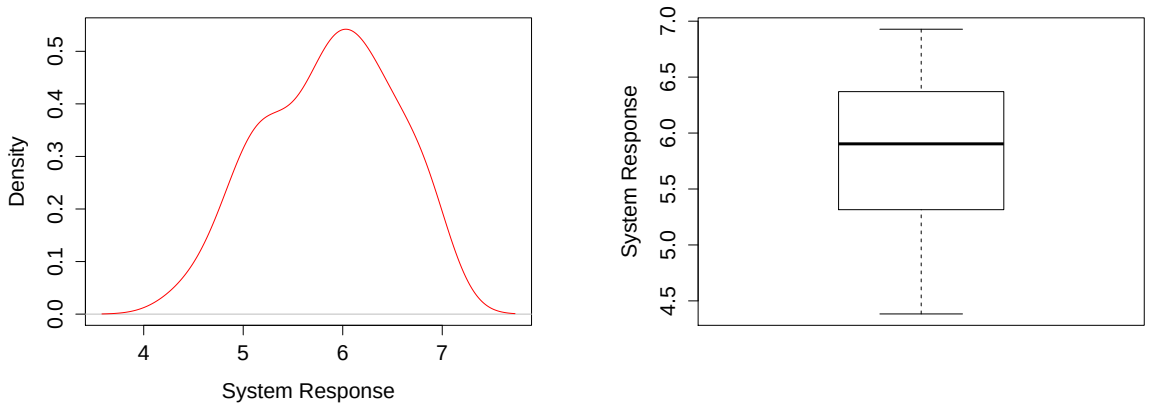
B. Optimized model selection procedure

In order to fine-tune the three metamodels, we applied different approaches to each.

For the polynomial model, starting from Eq. (8), we used a stepwise selection procedure based on the Akaike Information Criterion (AIC) [34] in order to identify the significant variables.

The resulting best model is:

$$Y = \beta_0 + \beta_1 x_2 + \beta_2 x_3 + \beta_3 x_2^2 + \beta_4 x_3^2 + \beta_5 x_1 x_2 + \epsilon, \tag{14}$$



(a)Density plot (b)Box plot

FIG. 5: Analysis of the system response

where x_1 is the voltage, x_2 the speed and x_3 the feed rate, β_i (with $i = 0, \ldots, 5$) are the regression coefficients and ϵ the random noise in the observable response, Y.

For the neural network and kriging models we applied a bootstrap procedure [35]. In the case of the neural network, the bootstrap procedure is used to identify the best number of nodes in the hidden layer and to set the decay parameter, λ. For the kriging model, it is used to select the correlation function family.

The bootstrap procedure works as follows: from a subset, of the initial data set (of size $n_t = 40$), the parameters of a range of different model specifications are estimated. Then, the different estimated models are used for predicting the system response associated with the remaining experimental points (n_v). At this point, the Mean Square Error (MSE) is calculated for each specification. The MSE is defined as:

$$MSE = \frac{1}{n_v} \sum_{i=1}^{n_v} \left(\hat{Y}_i - Y_i \right)^2, \tag{15}$$

where $\hat{Y}_i$ is the i^{th} predicted value of the system response. The procedure is repeated $B = 20$ times, each one with a different re-sampling of the initial subset n_t. Finally, the specification of each statistical model with the minimum averaged MSE is selected.

The best selected network topology (see Fig. 6) involves one hidden layer with seven neurons and a sigmoid activation function between the hidden layer and the output. The training algorithm for the weight selection is the back-propagation algorithm with optimized decay parameter $\lambda = 0.1$.

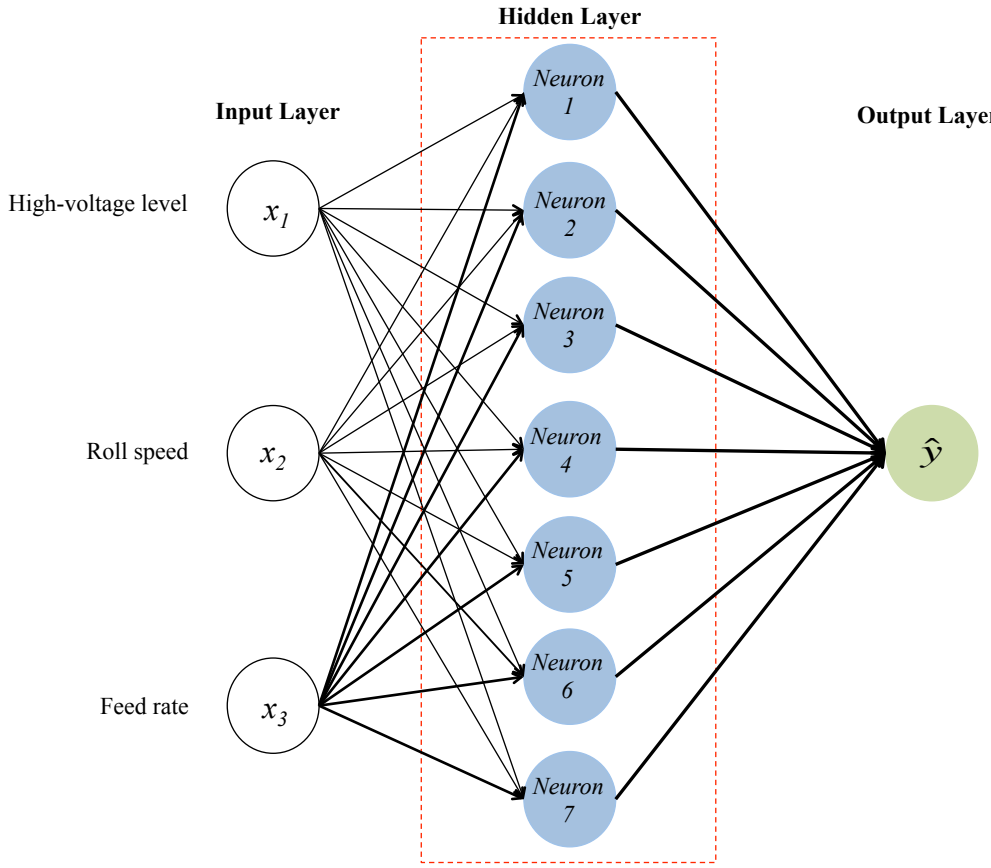


FIG. 6: Best feed-forward ANN with 7 neurons in the hidden layer. The different line thickness represents the importance of the connection arcs.

The best type of kriging model is based on the universal kriging approach and the selected most representative correlation function is the power exponential family which has been selected in accordance with minimum MSE obtained by applying the following correlation functions: power exponential ($MSE = 0.1037$), exponential ($MSE = 0.1145$), Matérn $\frac{3}{2}$ ($MSE = 0.1351$), Matérn $\frac{5}{2}$ ($MSE = 0.1551$) and power exponential with all parameters $p_i = 2$ ($MSE = 0.1394$), with $i = 1, \ldots, 3$.

C. Results

The three metamodels have been compared in terms of prediction accuracy by means of a cross-validation approach. In our case, we randomly divided $R = 50$ times the initial dataset into two subsets: a training set of size $n_t = 40$ and a validation set of size $n_v = 10$. Every time, the training set is used to estimate the coefficients and/or parameters

of the metamodels. The resulting metamodels are used to predict the system response of the validation set and then to evaluate their performance. The predictive performance indicators selected to compare the metamodels are: the mean squared error (MSE), the mean absolute error (MAE), the standard deviation of absolute error ($St.\ AE$), the mean absolute percentage error ($MAPE$) and the standard deviation of absolute percentage error ($St.\ APE$).

$$AE = \left|\hat{Y}_i - Y_i\right|, \qquad i = 1, \ldots, n_v \tag{16}$$

$$MAE = \frac{1}{n_v}\sum_{i=1}^{n_v} AE_i, \tag{17}$$

$$St.\ AE = \sqrt{\frac{1}{n_v - 1}\sum_{i=1}^{n_v}\Big(AE_i - MAE\Big)^2} \tag{18}$$

$$APE = \left|\frac{\hat{Y}_i - Y_i}{Y_i}\right|, \qquad i = 1, \ldots, n_v \tag{19}$$

$$MAPE = \frac{1}{n_v}\sum_{i=1}^{n_v} APE_i, \tag{20}$$

$$St.\ APE = \sqrt{\frac{1}{n_v - 1}\sum_{i=1}^{n_v}\Big(APE_i - MAPE\Big)^2} \tag{21}$$

where $\hat{Y}_i$ is the i^{th} predicted value of the response system in the r^{th} re-sample.

In Table I, the averages of the performance indicators, over $R = 50$ runs, are reported. From these results, the neural network shows better performance in terms of prediction accuracy. Besides that, the ability of the kriging model to predict the unknown system responses is comparable with the neural network. The selected polynomial model seems to perform worse but still acceptably. In Fig. 7, the MSE distribution is shown.

Given these results, we estimated the three metamodels on the $n = 50$ available data points selected by the LHS technique (see Sec. IV A).

In Fig. 8, 9 and 10, we show the different predicted response surfaces by means of contour plots. Each contour plot represent the behavior of the response system varying two of the three variables. The third variable is just kept fixed at the central value of its range.

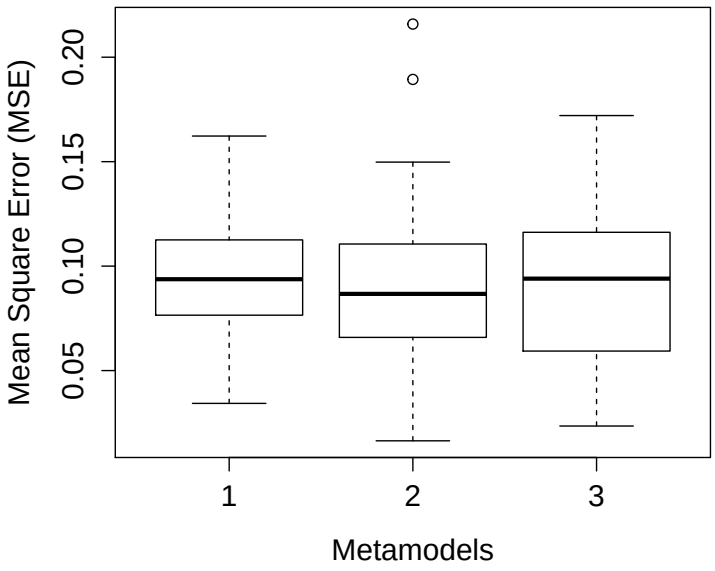


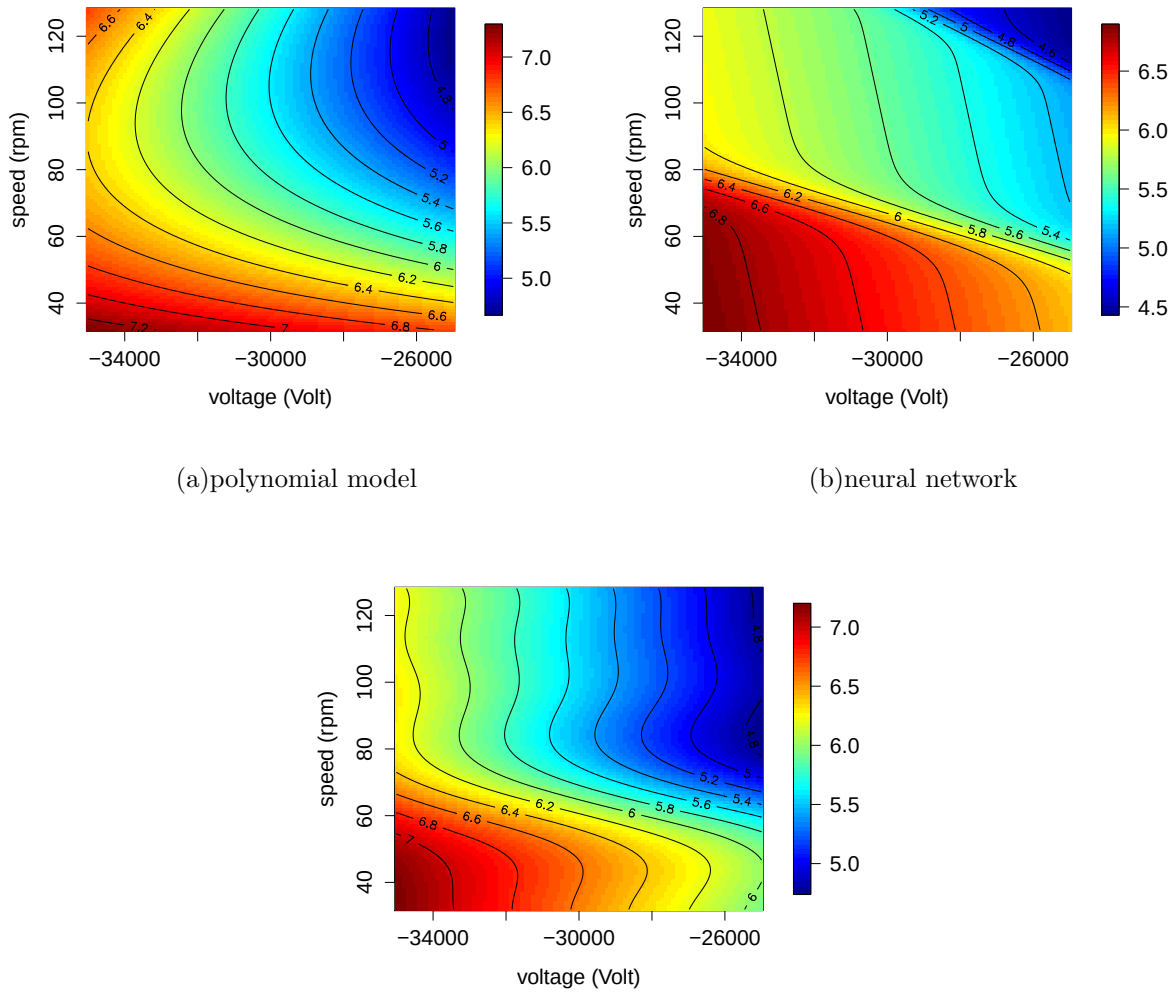
FIG. 7: Comparison of the MSE distribution for polynomial model (1), neural network (2) and kriging model (3).

	polynomial model	**neural network**	**kriging model**
MSE	0.0940	**0.0906**	0.0911
MAE	0.6421	**0.6034**	0.6063
St. AE	0.7618	**0.7291**	0.7332
MAPE	0.1113	0.1056	**0.1055**
St. APE	0.1325	0.1285	**0.1282**

TABLE I: Average values of the prediction performance indicators for the three metamodels.

The three metamodels give different levels of smoothness. In Fig. 8(a), the polynomial model identifies a concave surface with a maximum response value in the region of the minimum values for voltage and drum speed. The maximum is found in the same region by the other two response surfaces obtained with the Neural Network (Fig. 8(b)) and the kriging model (Fig. 8(c)), respectively. This result can be explained using forces balance. More specifically, by increasing the magnitude of the voltage, the electric force acting on booth particles is also increased, pushing metals far from the drum and attracting non-

metals on the drum surface. Reducing the drum speed, the contribution of the centrifugal force, which is a significant term, is reduced too, making the effects of the electrostatic forces more evident.



(a)polynomial model

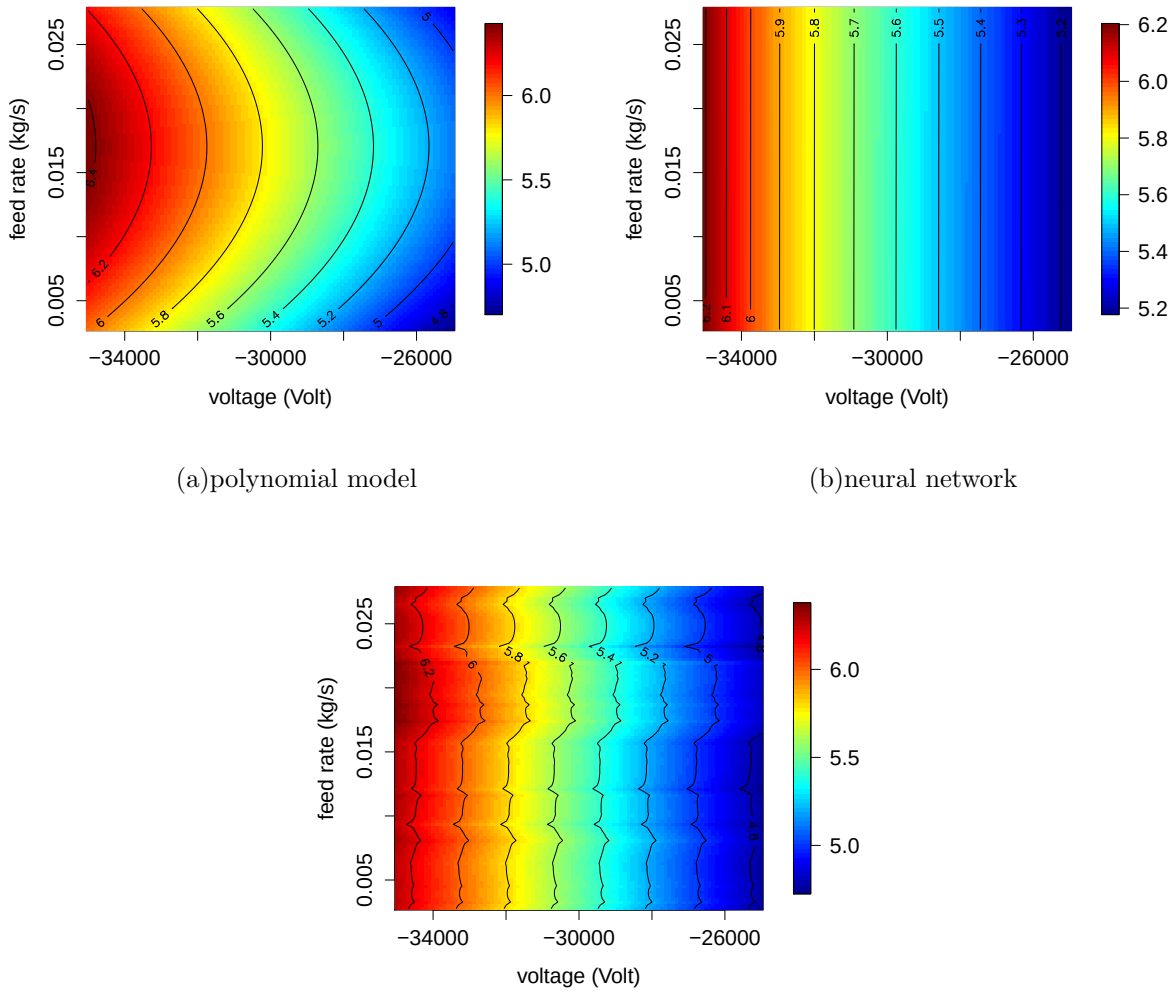
(b)neural network

(c)kriging model

FIG. 8: Contour plots of voltage and drum speed, or simply speed.

In Fig. 9 the response surfaces as functions of feed rate and voltage are considered. The polynomial model (Fig. 9(a)) is still a concave surface which leads to a specific region of optimality. In other words, the feed rate, given minimum values of voltage, influences the performance of the system. The neural network (Fig. 9(b)) identifies a slope surface with maximum values when voltage is set at the minimum, and minimum values when voltage is set at the maximum. In this case, the response surface is not affected by variations of the

feed rate. The same situation occurs with the kriging model (Fig. 9(c)) which identifies less straight contour lines but with the same feature.



(a)polynomial model

(b)neural network

(c)kriging model

FIG. 9: Contour plots of voltage and feed rate.

In Fig. 10, drum speed and feed rate are studied. In Fig. 10(a) (polynomial model), a stationary point is identified between 90 and 110 rpm and 0.015 and 0.018 kg/s. From 65 to 32 rpm the response system constantly increases towards maximum values when the feed rate is between 0.015 and 0.020 kg/s. Instead, the neural network (Fig. 10(b)) considers the feed rate non influential for the behavior of the response system and a slope between 50 and 75 rpm is identified. A similar slope is identified by the kriging model (Fig. 10(c)). However, the feed rate affects the response system and there is also interaction between drum speed

and feed rate, a feature absent from the other models. In fact, given minimum values of speed, maximum values of response system are reached in correspondence of higher values of feed rate.

The weak dependence of the response on the feed rate suggests two main remarks. Probably, the effect on the distribution of particles in the simulated flows, and consequently on the objective function, is noticeable for very high values of the feed rate. These values have not been tested. In fact, the experiments were carried out in the range expected from the normal use of the machine. In addition, it is not possible to generalize this behavior of the separation process because it is strictly linked to the specific properties of the material used in simulations. The problem needs to be studied in greater depth, for a better understanding of the physics of granular flows in terms of feed rate interaction with flow variability.

The three metamodels are slightly different in terms of predicted response surfaces. All of them have individuated a similar behavior of the predicted response as a function of voltage and drum speed but, concerning the influence of the feed rate, they have predicted different situations.

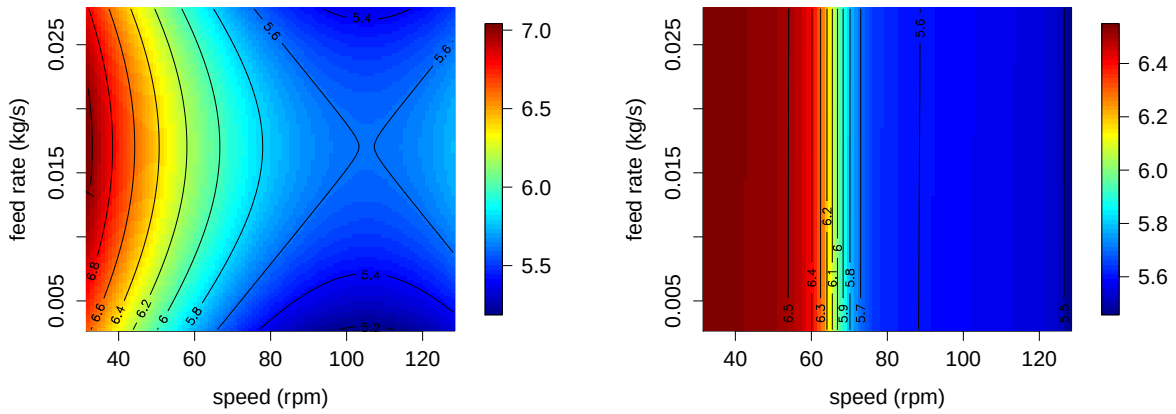
In order to understand which of the three metamodels better locates the optimal controllable parameters combination, we applied a Simulated Annealing (SA) [36, 37] optimization algorithm.

We ran the SA over the three metamodels obtaining identical results for the neural network and the kriging models. The optimal setting of the controllable parameters found in the polynomial response surface reflects the previous considerations and a different best value for the feed rate has been selected.

Subsequently, the CES simulator was used to calculate the response of the all optimal settings. Results reported in Table II show that the response from the simulated model is the same. This suggests that the feed rate does not affect the response system, given the minimum values of voltage and speed, as the neural network and the kriging model were correctly predicting.

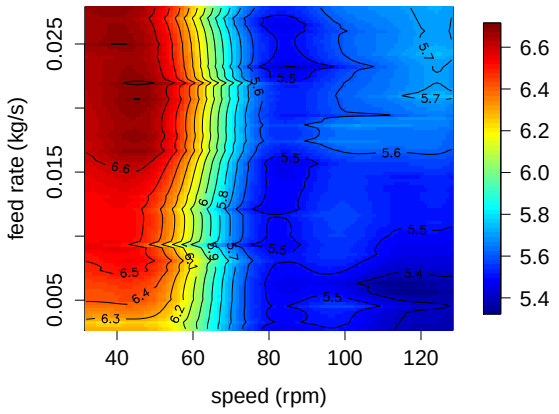
D. Kriging model Vs Neural Networks

Artificial Neural Networks (ANNs) are well-known models which have exhibited excellent behavior in the resolution of complex problems. It is demonstrated that feedforward ANNs



(a)polynomial model

(b)neural network



(c)kriging model

FIG. 10: Contour plots of drum speed, or simply speed, and feed rate.

are universal approximators, which means that an ANN can approximate, to any desired degree of accuracy, any real-valued continuous function [38].

Despite the success of ANNs, one of the most common criticism of ANNs is that there is no satisfactory interpretation of their behavior because they capture relations between inputs and outputs with a highly accurate approximation, but no definitive answer has been given about their inner workings.

Two important issues arise while applying modeling techniques to real applications: select the most powerful modeling approach (e.g. in terms of prediction) and better understand the problem under study. A large number of scientific papers are focused on comparing ANNs

	Optimal setting	**Predicted response**	**Simulated response**
polynomial model	-35000 | 32 | 0.01704	7.304	6.925
neural network	-35000 | 32 | 0.02778	6.899	6.925
kriging model	-35000 | 32 | 0.02778	7.312	6.925

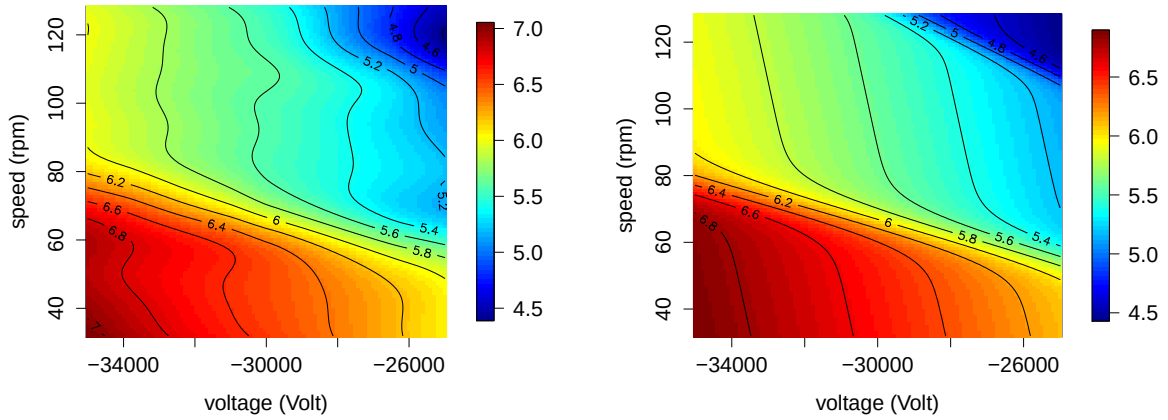
TABLE II: Optimal settings for voltage, speed and feed rate.

and kriging models [39–42], but we have found only a few theoretical works showing their equivalence under certain conditions. For example, R. M. Neal (1996) [43] demonstrated that some Bayesian regression models based on neural networks converge to a Gaussian processes as the number of neurons in the hidden layer tends to infinity. This result has motivated the use of Gaussian process models to problems which are typically addressed applying neural networks, such as high-dimensional applications [44] or classification problems [45].

Furthermore, Benítez et al. (1997) [46] demonstrated the equivalence between fuzzy rule-based systems (FRBS), developed using fuzzy logic, and ANNs. In their work it is shown that ANNs can be encoded by a continuous FRBS.

Following a similar reasoning, we show that it is possible to empirically obtain an *equivalence-by-approximation* between kriging models and ANNs. In other words, given a simple single layer feed-forward neural network with a sigmoid activation function for the hidden neurons and a linear one for the output neurons, there exists at least one universal kriging model with power exponential correlation function which approximates the same real function in a way similar to the neural network. This is possible by tuning the θ and p parameters of the power exponential correlation function and the trend coefficients. In Fig. 11, it is shown that a universal kriging model, correctly tuned, can approximate the response surface predicted by the ANN fitted to our simulator output. In particular, the kriging model has been estimated using the ANN predicted values of the first batch of data ($n = 50$) selected by the LHS as response.

This approximation provides a handle to interpret the artificial neural network fitted on our CES data. In fact, θ and p parameters have a clear influence on the power exponential correlation function [47] and, consequently, on the universal kriging model. Furthermore, as the two metamodels have shown a similar performance while optimizing the controllable parameters combinations, this approximation is important for understanding whether a



(a)approximated neural network by kriging model

(b)neural network

FIG. 11: Comparison between neural network response surface and the one approximated by the kriging model.

different behavior in the whole experimental region reflects limitations of the metamodels themselves or actual characteristics of the simulator.

In Table III, all the parameters of the models are reported. Model 1 collects the trend coefficients, scale and power parameters and variance of the universal kriging model estimated on the LHS design and Model 2 collects the same information for the universal kriging model that approximates the single layer feed-forward neural network. Clearly, the variance parameter of Model 2 is smaller, because the surface predicted by the neural network is smoother than the original data.

An examination of the trend coefficients shows that the coefficient of the feed rate is reduced by two orders of magnitude in Model 2, compared to Model 1, whereas the remaining coefficients show little variation. If we plotted the correlation function in the direction of the feed rate we would observe that the correlation of Model 1 decreases rather more sharply than that of Model 2 and that the former is dominated by the latter over the operating region, implying a greater degree of smoothness of Model 2. This indicates that Model 2 estimate of the feed rate effect on the predicted response depends mostly on a very weak linear trend, with a high degree of smoothing (correlation is always above 0.75). On the other hand, Model 1 captures the effect of the feed rate through the trend component, while

TABLE III: Estimated trend coefficients, scale and power parameters and variances of the two universal kriging models. Model 1 is the universal kriging model estimated on the first batch of data ($n = 50$) selected by the LHS and Model 2 is the approximation of the artificial neural network by means of the universal kriging model.

	Trend coefficients				**Scale parameters**			**Power parameters**			**Variance**
	β_0	β_1	β_2	β_3	θ_1	θ_2	θ_3	p_1	p_2	p_3	σ^2
Model 1	2.503	-0.0001	-0.012	10.151	12612.404	18.796	0.049	2.000	2.000	0.117	0.128
Model 2	3.583	-0.0001	-0.014	0.381	12409.871	12.597	0.049	1.186	1.994	2.000	0.048

keeping the long-term correlation above 0.4, still allowing for correlation between distant experimental points. For both models, according to the trend component, the influence of voltage is one order of magnitude greater than the influence of speed and the correlation function of Model 1 dominates that of Model 2 in both directions, but without a substantial difference.

This analysis leads to some important conclusions. First of all, an ANN may be expressed in a more comprehensible way by means of a universal kriging model. Secondly, from the application point of view, both the kriging model and the ANN have encoded a greater sensitivity of the response to voltage, which can be crucial for the real CES process, because small variations over the dielectric strength can cause an electrical discharge (by electric arch) and, consequently, they can alter the electric field, thus affecting the quality of output. Thirdly, the kriging model seems preferable to the ANN model, because it is capable of giving some more information on the effect of the feed rate.

Finally, it is common to consider two sources of uncertainty associated with ANN outcomes: uncertainty in training dataset and uncertainty in model structure [48]. Given these uncertainty sources, it could be informative to explore the prediction error of ANN, and to quantify how these sources might affect the prediction. Different studies [48, 49] have explored the behavior of ANN, investigating how to build confidence intervals and prediction intervals. However, selection and application of construction methods for exploring uncertainty in ANN depend on the purpose of the analysis and on computational constraints. Given these limitations, it is not always possible to calculate confidence intervals and predic-

tion intervals especially in real applications. A similar difficulty occurs in kriging modeling when the form of the correlation function is not known. In this case, computational and empirical techniques are recommended. A different situation arises when $Z(\mathbf{x})$ is a Gaussian stationary process with $E(Z(\mathbf{x})) = 0$ and constant variance σ_Z^2 and the form of correlation function $R(\cdot|\psi)$ is known, depending on some parameters ψ. Under these hypotheses, an unbiased predictor is obtained by minimizing the Mean Squared Prediction Error (MSPE) and a closed form of prediction intervals is available. When ψ is unknown it must be estimated, and the work of Santner et al. (2003) [2] provides ways of taking the estimation uncertainty into account.

V. CONCLUSIONS

The rapid growth of waste electrical and electronic equipments (WEEE) has led to an increasing demand from experimental fields for efficient methods to improve material separation processes, such as the corona electrostatic separation.

This work has addressed this issue by proposing a computer-aided methodology, derived by computer experiments, and comparing three different modeling techniques: polynomial regression, kriging and artificial neural network (ANN) models. Our numerical results in Section IV C suggest that kriging and ANN models are equivalent in terms of optimization. In fact, the optimal setting of the controllable parameters (voltage, drum speed and feed rate) obtained by the two metamodels is the same. Furthermore, both models show the importance of voltage, because its influence on the response is one order of magnitude greater than the influence of speed. Where the kriging and the ANN models mainly differ is in how they treat the feed rate. On the one hand, ANN considers the effect of feed rate on the predicted response as weakly linear and as not-interacting with voltage or drum speed. On the other hand, the kriging model reveals the effect of larger feed rates on the predicted response. Therefore, considering the known physical properties of CES, the most suitable metamodel seems to be the kriging model, which better describes some properties of the CES process.

Another interesting finding of our analysis of the two metamodels is that it is possible to empirically obtain a *equivalence-by-approximation* between the kriging and the ANN models. In fact, by fine tuning the θ and p parameters of the power exponential correlation function

and the trend coefficients in the kriging model, an approximation of the response surface close to that predicted by an ANN can be obtained, which can aid interpretation of the ANN fit.

With the objective of optimizing the response surface, further studies can be done in the direction of sequential experimental design approaches. An interesting starting point could be the work of Schonlau et al. (1998) [50] which is based on Bayesian optimization for statistical models, typically stochastic processes, capable of handling highly nonlinear relationships. The idea is that, using a sequential approach, after few evaluations of the computer model (i.e. the CES simulator) one can improve the quality of the statistical model and optimize the simulator using specific optimal criteria, such as the expected improvement criterion.

Finally, we remark that our computer-aided methodology can be successfully applied to CES to handle experiments with a higher number of process variables, with a different level of accuracy, based on the chosen metamodel. More generally, we believe that further exploration in this direction may generate valuable contributions to the optimization and description of the recycling process of WEEE, because also other stages of this process, taking place before separation, require modeling and optimization.

Acknowledgments

This work has been carried out as a part of the FIDEAS project (Fabbrica Intelligente per la Deproduzione Avanzata e Sostenibile) co-funded within the Framework Agreement between Regione Lombardia and CNR. The authors would like to thank the two anonymous referees for their suggestions which helped to improve the quality of the work.

---

[1] Montgomery, D. C., *Design and analysis of experiments*, Wiley, 2004.

[2] Santner, T. J., Williams, B. J., and Notz, W. I., *The design and analysis of computer experiments*, Springer, 2003.

[3] Alvarez, M. J., Gil-Negrete, N., Ilzarbel, L., Tanco, M., Viles, E., and Asensio, A., A computer experiment application to the design and optimization of a capacitive accelerometer, *Appl. Stochastic Models Bus. Ind.*, Vol. 25, pp. 151-162, 2009.

[4] Fang, K.T., Li, R., and Sudjianto, A., *Design and Modelling for Computer Experiments*, Chapman and Hall/CRC, London, 2006.

[5] Berni, R., De March, D., and Stefanini, F. M., T-optimality and neural networks: a comparison of approaches for building experimental designs, *Appl. Stochastic Models Bus. Ind.*, Vol. 29, pp. 454-467, 2013.

[6] Pedone, P., Vicario, G., Romano, D., kriging-based sequential inspection plans for coordinate measuring machines, *Appl. Stochastic Models Bus. Ind.*, Vol. 29, pp. 133-149, 2009.

[7] Baldi Antognini, A., Giovagnoli, A., Romano, D., and Zagoraiou, M., Computer simulations for the optimization of technological processes, In Erto P. (Ed.), *Statistics for Innovation*, 65-86, 2009.

[8] Ralston, O. C., *Electrostatic separation of mixed granular solids*, Elsevier, 1961.

[9] Kiewiet, C., Bergougnou, M. A., and Brown, J. D., Electrostatic separation of fine particle in vibrate fluidized beds, *IEEE Trans. Ind. Appl.*, Vol. 6, pp. 526-530, 1978.

[10] Taylor, J. B., Dry electrostatic separation of granular materials, *Proceeding of the IEEE IAS Annual Meeting*, Vol. 35, pp. 1741-1759, 1988.

[11] Grübler, A., *Technology and global change*, Cambridge University Press, 2003.

[12] Dascalescu, L., Tilmatine, A., Aman, F., and Mihailescu, M., Optimization of electrostatic separation processes using response surface modeling, *IEEE Trans. Ind. Appl.*, Vol. 40, pp. 53-59, 2004.

[13] Hou, S., Wu, J., Qin, Y., and Xu, Z., Electrostatic separation for recycling waste printed circuit board: a study on external factor and a robust design for optimization, *Environ. Sci. Technol.*, Vol. 44, pp. 5177-5181, 2010.

[14] Mihailescu, M., Samuila, A., Urs, A., Morar, R., Iuga, A., and Dascalescu, L., Computer-assisted experimental design for the optimization of electrostatic separation processes, *IEEE Trans. Ind. Appl.*, Vol. 38, pp. 1174-1181, 2002.

[15] Touhami, S., Medles, K., Dahou, O., Tilmatine, A., Bendaoud, A., and Dascalescu, L., Modelin and optimization of a roll-type electrostatic separation process using artificial neural netwroks, *IEEE Trans. Ind. Appl.*, Vol. 49, pp. 1773-1780, 2013.

[16] Critelli, I., Degiorgi, A., Colledani, M., and Tasora, A., A simulation model of corona electrostatic separation (CES) for the recycling of printed circuit boards (PCBs), *Proceeding of SUM2014*, Vol. 2, pp. 1-20, 2014.

[17] Li, J., Lu, H., Xu, Z., and Zhou, Y., A model for computing the trajectories of the conducting particles from waste printed circuit boards in corona electrostatic separators, *Journal of Hazardous Materials*, Vol. 151, pp. 52-57, 2008.

[18] Younes, M. , Tilmatine, A., Medles, K., Rahli, M., and Dascalescu, L., Numerical modeling of conductive particle trajectories in roll-type corona-electrostatic separators, *IEEE Transactions on Industry Applications*, Vol. 43, pp. 1130-1136, 2007.

[19] Critelli, I., Tasora, A., Degiorgi, A., and Colledani, M., Particle simulation of granular flows in electrostatic separation processes, *Proceeding of SIMUL 2014 Sixth International Conference on Advances in System Simulation*, October 2014.

[20] Anitescu, M., and Tasora, A., An iterative approach for cone complementarity problems for non smooth dynamics, *Computational Optimization and Applications*, Vol. 47, pp. 207-235, 2010.

[21] Allenby, B. R., and Richards, D. J., *The greening of industrial ecosystems*, National Academy Press, 1994.

[22] Gutowski, T. G., Wolf, M. I., Dahmus, J. B., and Albino, D. K., Analysis of recycling systems, *Proceedings of 2008 NSF Engineering Research and Innovation Conference*, pp. 1-8, 2008.

[23] Myers, R. H., Montgomery, D. C., and Anderson-Cook, C. M., *Response surface methodology: process and product optimization using designed experiments*, Wiley, 2009.

[24] Ripley, B. D., *Pattern recognition and neural networks*, Cambridge University Press, 1996.

[25] Rumelhart, D. E., Hinton, G.E., and Williams, R.J., Learning internal representations by error propagation, In Rumelhart, D. E., and McClelland, J.L. (Eds.), *Parallel Distributed Processing*, MIT Press, pp. 318-362, 1986

[26] Karlik, B., and Olgac, A. V., Performance analysis of various activation functions in generalized MLP architectures of neural networks, *International Journal of Artificial Intelligence And Expert Systems*, Vol. 1, pp. 111-122, 2010.

[27] Krige, D.G., A statistical approach to some basic mine valuation problems on the Witwatersrand, *J. of Chem., Metal. and Mining Soc. of South Africa*, Vol. 52, pp. 119-139, 1951.

[28] Lebensztajn, L., Marretto, C.A.R., Costa, M.C., and Coulomb, J. L., Kriging: a useful tool for electromagnetic device optimization, *IEEE Transactions on Magnetics*, Vol. 40, pp. 1196-1199, 2004.

[29] Pan, I., and Das, S., Kriging based surrogate modeling for fractional order control of micro-

grids, *IEEE Transactions on Smart Grid*, Vol. 6, pp. 36-44, 2015.

[30] Sacks, J., Welch, W., Mitchell, T.J., and Wynn, H.P., Design and analysis of computer experiments, *Stat. Sci.*, Vol. 4, pp. 409435, 1989.

[31] Simpson, T. W., Lin, D. K. J., and Chen, W., Sampling strategies for computer experiments: design and analysis, *International Journal of Reliability and Application*, Vol. 2, pp. 209-240, 2001.

[32] McKay, M.D., Beckman, R.J., and Conover, W.J., A comparison of three methods for selecting values of input variables in the analysis of output from a computer code, *Technometrics*, Vol. 21, pp. 239-245, 1979.

[33] Stocki, R., A method to improve design reliability using optimal Latin hypercube sampling, *Computer Assisted Mechanics and Engineering Sciences*, Vol. 12, pp. 87-105, 2005.

[34] Davison, A.C., *Statistical models*, Cambridge University Press, 2003.

[35] De March, D., Borrotti, M., Sartore, L., Slanzi, D., Podestá, L., and Poli, I., A predictive approach based on neural network models for building automation systems, In Bassis, S., Esposito, A., and Morabito, F. C. (Eds.), *Advances in Neural Networks: Computational and Theoretical Issues*, Springer, Vol. 37, pp. 253-262, 2015.

[36] Kirkpatrick, S., Gelatt, C. D., and Vecchi, M. P., Optimization by simulated annealing, *Science*, Vol. 220, pp. 671-680, 1983.

[37] Cerny, V., Thermo dynamical approach to the travelling salesman problem: an efficient simulation algorithm, *J. Optim. Theory Appl.*, Vol. 45, 41-51, 1985.

[38] Hornik, K., Stinchcombe, M., and White, H., Multilayer feedforward networks are universal approximators, *Neural Networks*, Vol. 2, 359-366, 1989.

[39] Villa-Vialaneix, N., Follador, M., Ratto, M., and Leip, A., A comparison of eight metamodeling techniques for the simulation of N_2O fluxes and N leaching from corn crops, *Environmental Modelling and Software*, Vol. 34, pp. 51-66, 2012.

[40] Razavi, S., Tolson, B. A., and Burn, D. H., Numerical assessment of metamodelling strategies in computationally intensive optimization, *Environmental Modelling and Software*, Vol. 34, pp. 67-86, 2012.

[41] Luo, J., and Lu, W., Comparison of surrogate models with different methods in groundwater remediation process, *J. Earth Syst. Sci.*, Vol. 7, pp. 1579-1589, 2014.

[42] Jiang, C., Zhou, C., Liu, J., Lan, T., Yang, G., Zhao, Z., Zhu, P., Sun, H., and Cui, X.,

Comparison of the Kriging and neural network methods for modeling *foF2* maps over North China region, *Advances in Space Research*, 2015, available on-line.

[43] Neal, R. M., *Bayesian learning for neural networks*, Lecture Notes in Statistics, Springer, 1996.

[44] Williams, C. K. I., and Rasmussen, C. E., Gaussian processes for regression, In Touretzky, D. S., Mozer, M. C., and Hasslemo, M .E. (Eds), *Advances in Neural Information Processing Systems*, MIT Press, Vol. 8, pp. 514-520, 1996

[45] Neal, R. M.,Monte Carlo implementation of Gaussian process models for Bayesian regression and classification, *arXiv:physics/9701026v2*, 1997.

[46] Benítez, J. M., Castro, J. L., and Requena, I., Are artificial neural networks black boxes?, *IEEE Transaction on Neural Network*, Vol. 8, 1156-1164, 1997.

[47] Stein, M., *Interpolation of spatial data: some theory of kriging*, Springer, New York, 1999.

[48] Mazloumi, E., Rose, G., Currie, G., and Moridpour, S., Prediction intervals to account for uncertainties in neural network predictions: Methodology and application in bus travel time prediction, *Engineering Applications of Artificial Intelligence*, Vol. 24, pp. 534-542, 2011.

[49] Khosravi, A., Nahavandi, S., Creighton, D., and Atiya, A. F., Comprehensive review of neural network-based prediction intervals and new advances, *IEEE Transactions on Neural Network*, vol. 22, pp. 1341-1356, 2011.

[50] Schonlau, M., Welch, W. J., and and Jones, D. R., Global versus local search in constrained optimization of computer models, In Flournoy, N., Rosenberger, W. J., and Wong, W. K. (Eds.), *New Developments and Applications in Experimental Designs*, IMS Lecture Notes-Monograph Series, Vol. 34, pp. 11-31, 1998.