

Hybrid model for improving width performances in hot strip mill

J.-L. Rendueles, V. Lobato (Arcelor-Mittal, Avilés)
F. Ortega (Oviedo University)
M.-T. Rodriguez (Knowledge Innovation RC, Arcelor Mittal)

A modelling strategy has been developed to reduce the overspread at the finisher of the HSM in Avilés. A new hybrid data based model elaborates the roughing and finishing mills set up. It has made it possible to reduce the overspread by 1.5 mm at the finisher since March 2005. The model is modular and flexible and it can be transferred to other installations.

Subject of a presentation at the 9th International, 4th European Steel Rolling 2006 Conference (ATS, Paris, June 19-21, 2006)

■ INTRODUCTION

The control of the dimensional factors in steel making is becoming an increasingly complex task because of the strict requirements for high productivity, low energy consumption, and production of different high quality steel grades.

Accurate prediction of the width deformations in the roughing mill and the spread in finishing mill during hot rolling is important for both transverse and thickness dimensional control. Width deformation is usually estimated by means of mathematical models and repetitive control techniques. This approach provides good results for normal conditions but has drawbacks when the type of material is changed or when modifications in the rolling processes are done. For these cases is quite difficult, first to identify when it is necessary to adapt the model, and second to adapt it in a proper way.

This paper presents the strategy and results of the research developed by Arcelor Mittal in collaboration with the University of Oviedo to reduce the overspread obtained at the output of the Hot Strip Mill (HSM) at the Avilés plant (Spain). The modelling strategy that has been developed has taken advantage of the existing model operating in HSM and a new hybrid data based model developed with APIMARS technique (4). The mathematical model calculates a preset as accurate as possible, while the hybrid data based model, by using the output of mathematical model as input together with other process variables, elaborates the final roughing and finishing mills setup.

Mathematical models are usually based on geometrical considerations, as relations among width, contact length, initial thickness, rolls diameters, and many other mill configurations. Also, several lookup tables are elaborated and used by mathematical models giving at the end linearized results. With the approach here explained the non linearities are also covered and substantial improvement is clearly reached.

The techniques used for the creation of the models are first described. Then, it is explained how the techniques were combined and the strategies defined for modelling. The data used during the research and the variables detected as relevant for the control of the overspread are further reported. The results obtained during off-line tests and since the on-line integration of the models in March of 2005 are also presented.

■ DESCRIPTION OF THE TECHNIQUES EMPLOYED

Overall, the techniques that have been used to improve control of the mill spread falls within the so-called "Data Mining" type. This is the name given to an interdisciplinary

Modèle hybride pour amélioration des performances largeur au train à bandes

J.-L. Rendueles, V. Lobato (Arcelor-Mittal, Avilés)

F. Ortega (Oviedo University)

M.-T. Rodriguez (Knowledge Innovation RC, Arcelor Mittal)

La prévision de l'élargissement au finisseur est essentielle pour l'amélioration de la performance dimensionnelle en épaisseur et en largeur au train à bandes. Le contrôle de l'élargissement est classiquement fondé sur des modèles mathématiques étayés par des mesures sur produits. Cette approche peut être prise en défaut par un changement de comportement de l'acier ou du schéma de laminage.

Une stratégie de modélisation a donc été développée pour diminuer les sur-largeurs au train à bandes d'Avilés. Un nouveau modèle hybride à base de données calcule les réglages du dégrossisseur et du finisseur. Cela a permis de réduire les sur-largeurs de 1,5 mm au finisseur depuis mars 2005. Le modèle est modulaire et adaptable ; il peut être transféré à d'autres installations.

Les techniques utilisées sont du type « data mining », avec des techniques de statistiques multi variables, des réseaux auto adaptatifs et l'algorithme Apimars de l'Université d'Oviedo qui permet la sélection semi automatique des paramètres pertinents et l'apprentissage des modèles prédictifs.

Le modèle hybride utilise les résultats d'un modèle classique et les mesures en ligne comme données sources, avec une procédure de sélection préliminaire.

Trois applications ont été développées, pour le dégrossisseur, pour le finisseur et pour l'ensemble dégrossisseur et finisseur.

Dans le cas du dégrossisseur, les résultats de la modélisation ont été sensiblement améliorés en distinguant les schémas de laminage à 5 ou à 7 passes. L'influence des teneurs en titane et en soufre de nouvelles nuances sur la dispersion de l'épaisseur a pu aussi être révélée par le modèle.

Dans le cas du finisseur, le modèle hybride a permis une amélioration très sensible du contrôle de l'élargissement, caractérisée par une diminution de la dispersion de l'ordre de 1 mm.

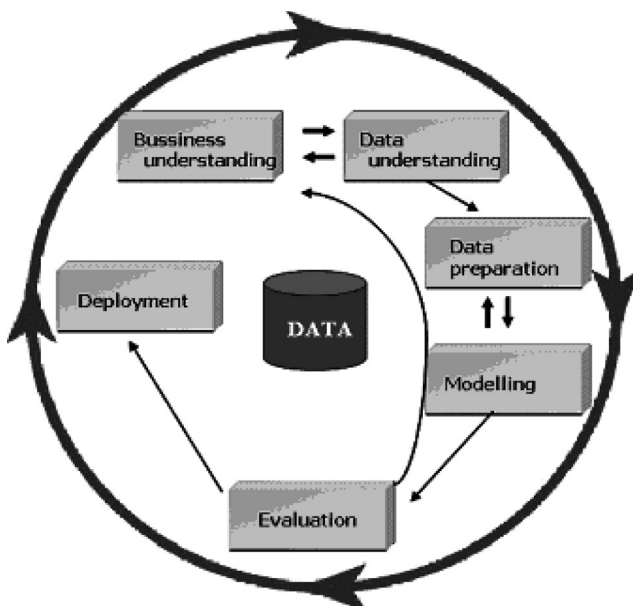


Fig. 1 - Cycle of life of the data mining projects based in the CRISP-DM methodology.

Fig. 1 - Séquence d'exécution des projets de data mining selon la méthode CRISP-DM.

field of techniques that are capable of extracting useful information from a simple piece of information inherent in the data. Data mining projects include several generic tasks, most of them repeated in an iterative and interactive way during the life cycle of projects. The CRISP-DM methodology (1) has been followed in order to develop the tasks in a structured manner. This methodology structures the data mining projects into six phases (fig. 1). The sequence of the phases is not rigid. Moving back and forth between different phases is always required. It depends on the outcome of each phase which phase or which particular task of a phase, has to be performed next. The arrows indicate the most important and frequent dependencies among phases.

Within the data-mining framework, the particular techniques used in this research were as follows:

- Multivariate statistics and display techniques: to carry out a preliminary analysis of the data and to become familiar with them. These techniques are normally used in the phases of data understanding and preparation, especially mentioned in the phases as variables selection and feature extraction. Both phases are very important for continuing the development of modelling, for instance outliers can cause bad modelling results and it is essential an initial treatment.

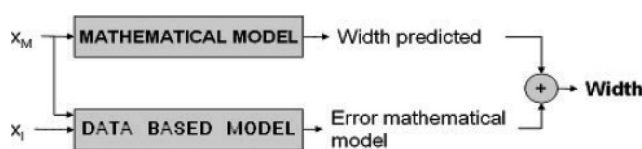


Fig. 2 - Original architecture, with data based model as corrector.

Fig. 2 - Architecture initiale, avec adaptation par le modèle à base de données.

- SOM Networks: also called self-organising maps (2), which have the capacity to group data according to their similarity. Moreover, the component drawings facilitate the display of multidimensional variables, extracting useful information about the relations among them.

- APIMARS: it is a modification introduced by the University of Oviedo to the MARS algorithm developed by Friedman in 1991 (3). APIMARS enables the semi-automatic selection of the most relevant variables and the training of predictive models.

■ DEFINITION OF STRATEGY

The project is based on the need to improve the restrictions of classic models. These models have been increasingly improved over recent years; however, they are still open to further improvement. The model chosen for this purpose were hybrid ones. *Figures 2 and 3* show the two types that have been tried out throughout the project. The structure indicated in *figure 3* was finally chosen on account of the better results obtained. This structure uses the classic model output as input for the hybrid model together with other signals coming from the process. All these signals are obtained from the preliminary selection process in accordance with their importance for the overall model during the business understanding and data preparation phases.

Due to fact that there are spread changes in both, roughing and finishing mills, the project developed three models:

- One to set up the roughing mill spread (RM),
- Another for the finishing mill spread (FM),
- An overall spread model, from entry of RM to FM output.

■ DATA SELECTION

The modelling is strictly based on data, thus the acquisition of them is of great importance. A big effort was made in this area for improving the acquisition system of the mill. Hardware and software were revamped for getting as many suitable variables as possible. A new computer was installed and a sophisticated database was developed under the frame of Oracle 9i®.

On the basis of the database created, 450 variables related to all process parameters were chosen, with a number of over 90,000 coils, trying to make the models fully representative, that means all possible case should be present in the data. For

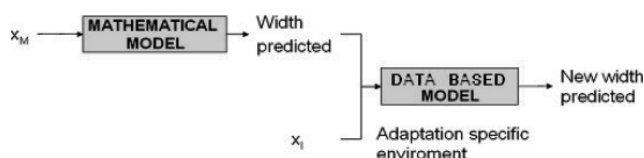


Fig. 3 - Current architecture, with the data based model as a generator of preset sent to the mill.

Fig. 3 - Architecture actuelle, avec préréglage du train par le modèle à base de données.

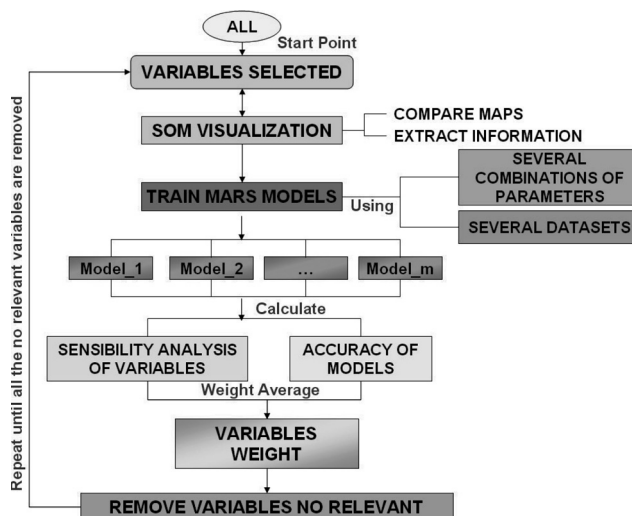


Fig. 4 - Scheme of the combined method of APIMARS-SOM for the selection of variables.

Fig. 4 - Principe de la méthode mixte APIMARS-SOM pour la sélection des variables.

the final selection of relevant variables, it was used a semi-automatic method developed by the University of Oviedo based in a wrapper strategy using APIMARS technique (4, 5).

APIMARS algorithm is a modification of the MARS algorithm developed by Friedman (3). MARS algorithm is capable to select automatically the variables intervening in the model for an input dataset. But if the input data set is slightly modified, for example adding or removing a case, a completely different set of variables selected can be obtained. To avoid this situation the strategy developed by the University of Oviedo was training several models using different datasets. The procedure followed is sketched in *figure 4*. For each model the variables are ranked using the lack of fit criterion. Results are merged using the average of the rankings weighted by the accuracy of the models. That is, the variables intervening in poor models are penalized. Variables not intervening in the models or with poor contribution are removed. The process is repeated until all the variables are identified as meaningful.

As an example of APIMARS, presentation of variables meaningful for a model of RM is shown in *figure 5*. Variables are ranked from 0 (less important) to 100 (more important). Ranking is the result of 10 models created using several datasets.

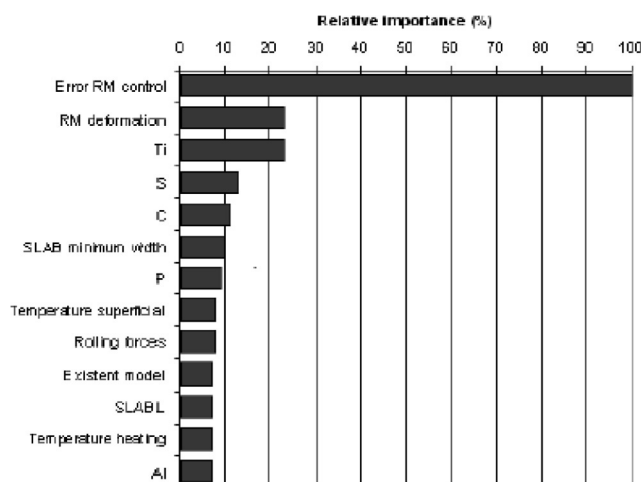


Fig. 5 - Variables detected as relevant RM adaptation (5 passes).

Fig. 5 - Détermination des paramètres pertinents au dégrossisseur (5passes).

Selection of variables following this method is blind: the expert knowledge is used in the initial selection of variables but the final selection is done in an automatic way.

During the selection of variables SOM networks (4) were incorporated to visualize the multidimensional space obtaining information relative to the connections among the variables or why variables could be relevant. This technique was very important for extracting knowledge from data and obtaining correlation among variables, removing in that way those variables given the same and redundant information.

The application of variables selection strategy provided very interesting information about some of the causes originating deviations in the width models, giving feedback to technicians of possible new phenomena that could allow the continuous improvement of process.

For example, it was detected that the width model did not considered in a proper way the influence of some chemical

elements such as titanium or sulphur (*fig. 6*) or the superficial temperature of the slabs. These results incorporated into the model led to a significant improvement.

RESULTS

Though many other models were developed, after on-line tests two were finally implemented:

- Spread at roughing output,
- Spread at finishing mill.

Roughing mill spread model

Several models were developed in order to predict spread, but the most favourable cases arose with separated models for 5 and 7 pass rolling strips. Spread deviations were taken as the difference between the average value measured by the gauges and the spread predicted by the mathematical model.

As mentioned above, the first trial was modelling a global model without distinguishing in number of passes. Results for this model were 88% of cases below a tolerance of 3 mm of error. These results were considered insufficient and after several estimations it was decided to separate the model by passes. The new model developed considering 5 or 7 passes enhanced the results. The new accuracy in off-line test was about 94% inside the tolerance of 3 mm of error. During the on-line tests of the last model, it was observed that these percentages remained constant and were considerably greater than those produced by the mathematical model, as can be seen in *table 1*.

These results become worst with the appearance of a new type of material that was not considered in previous cases. In the presence of such material, even the classic model behaved erratically. Several actions were taken following the CRISP_DM methodology again, and as consequence of these new studies it was discovered the relationship of sulphur (S) and titanium (Ti) content of the new material with the lateral flow behaviour (*fig. 6*).

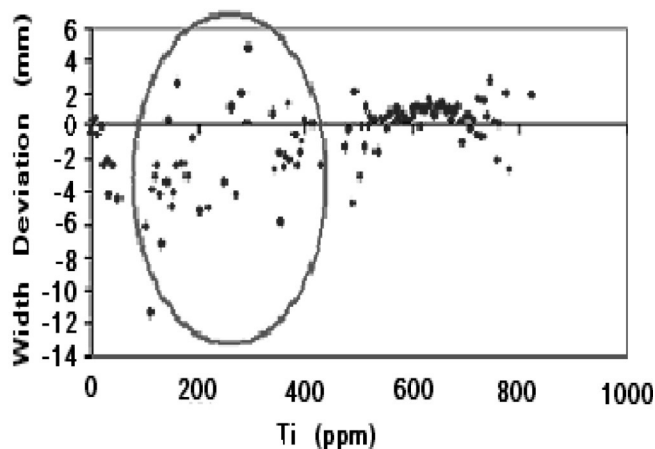
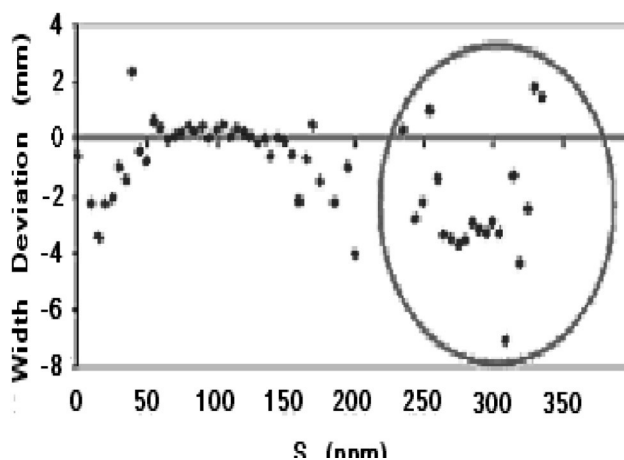


Fig. 6 - Sulphur and titanium levels as causes of spread deviation.

Fig. 6 - Dispersion de largeur en fonction des teneurs en soufre et en titane.

TABLE I: Comparative results obtained between data based and classic models during on-line tests.

TABLEAU I: Résultats comparés des modèles à base de données et des modèles classiques pour des essais en ligne.

Error (mm)	Mathematical model	Data based model
1	41%	59%
2	68%	86%
3	81%	94%
4	91%	98%
5	95%	99%
7	98%	99%
10	99%	100%

After elaborating again a new configuration of inputs for the model and retraining it again, the performance was newly back to similar values as before. In *figure 7*, it is possible to see such a new performance, between observations 25 to 45, both mathematical and initial hybrid models obtain bad output performances and big deviation error. The new one retrained overcame these difficulties giving similar performances as for other materials.

After introducing on-line, the results of width performance can be seen in *figure 8*, it is clear how the new hybrid model obtains better results than previous mathematical one existing in the roughing mill.

Anyway, even though the obtained results mean an improvement in roughing mill performance, they did not produce meaningful improvements at HSM delivery. Thus, a new model for finishing became necessary.

Modelling on finishing mill overspread

A model was created for the finishing mill overspread. The philosophy of working was conditioned by the way of working of the existing automation. In fact, all variables needed by the model are only ready some instants before the slab enters the RM. Thus, preset must be calculated considering this drawback, limiting the use of several variables.

Following again the same methodologies, the off-line results revealed the considerable improvement achieved by hybrid data based model as opposed to the existing mathematical one. In the *table II* it can be seen how about 93.8% of observation made are inside a tolerance of 3 mm and an improvement with regard to the finishing mill model of nearly 1 mm.

The on-line results were equally as good as initially expected. In *figure 9*, it is possible to see how the hybrid data based model adapts much better than the mathematical one. This figure shows several observations comparing real spread measured against both models. The hybrid model adapts much better to mill features that imply spread modifications.

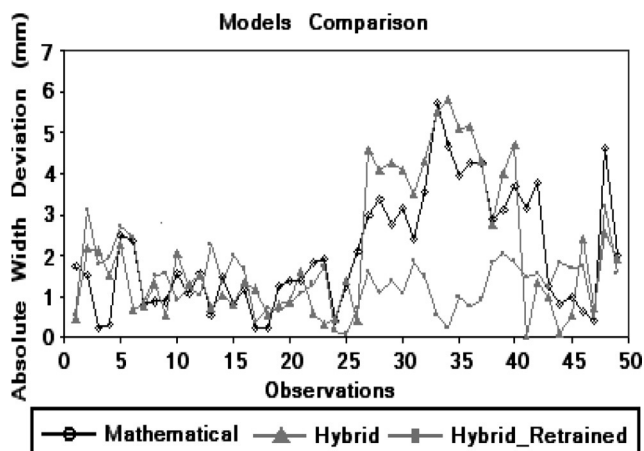


Fig. 7 - Comparison of the mathematical model, the hybrid model and the newly corrected hybrid model.

Fig. 7 - Comparaison du modèle mathématique, du modèle hybride et du modèle hybride récemment corrigé.

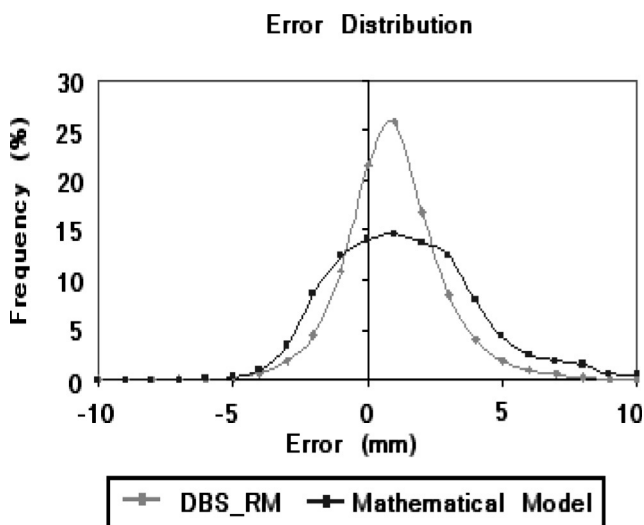


Fig. 8 - Comparison between final data based model (DBS_RM) and the mathematical model (FSU).

Fig. 8 - Comparaison entre le modèle à base de données final (DBS_RM) et le modèle mathématique (FSU).

TABLE II: Comparison of the off-line results between mathematical and data based models for the finisher.

TABLEAU II: Comparaison des résultats hors ligne au finisseur avec le modèle mathématique et le modèle à base de données.

Tolerance	Mathematical model	Data based model
1 mm	46.6%	54.0%
2 mm	72.1%	82.7%
3 mm	85.8%	93.8%
4 mm	92.4%	97.6%
5 mm	95.8%	98.9%

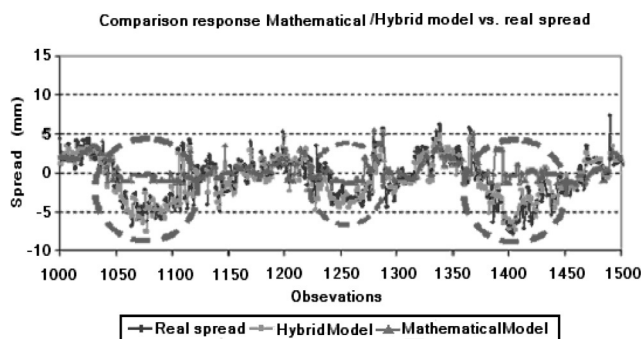


Fig. 9 - Comparison of the evolution of the overall spread measured with the mathematical and the data based models.

Fig. 9 - Comparaison de la dispersion d'élargissement avec le modèle mathématique et le modèle à base de données.

These results validated the implementation of the model, thus the hybrid has been working as the main mill spread set-up since 2005. As result, it has been achieved an accuracy improvement of nearly 1.5 mm.

Also, it was tested a new performance joining both models but results were considered not a significant improvement, though they were good results as well.

■ CONCLUSIONS

In accordance with the results it may be concluded that:

- The combination of the existing mathematical models and the hybrid data based model has enabled the generation of spread presets both on the roughing, as well as the finishing mill, improving the results of previous models about 1.5 mm. The improvement with regard to the previous year's values can be observed in *figure 10*.
- The structure of the model used enables the combination of classic and data based models, which results in a much more solid application.
- The system has been in operation practically at 100% capacity since March 2005, which proves its fitness and versatility after tens of thousands of proven cases.
- The models are modular and easy to update without touching the code. Moreover, they can be extrapolated conceptually, both to other installations as well as to other problems in the same installation.

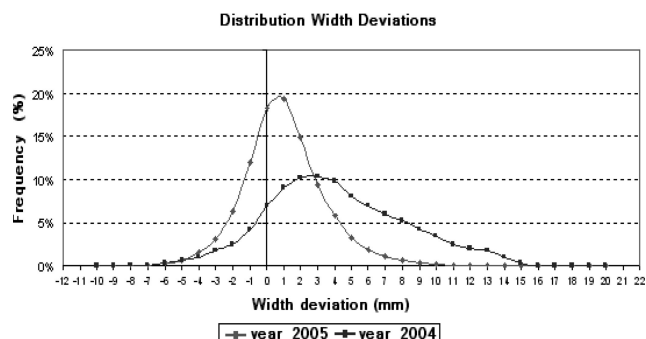


Fig. 10 - Spread improvement obtained as a result of the project.

Fig. 10 - Amélioration de l'élargissement obtenu grâce au projet.

- During the data mining process several studies of variables have been done giving feedback information to technical personal that was very useful for further improvement of HSM managing.

- Some problems were detected that it has been possible to solve as a result of the project. The case of the materials with unusual levels of S and Ti represents the most important example of such problems.

■ BIBLIOGRAPHY

- (1) (P.) CHAPMAN, (J.) CLINTON, (R.) KERBER, (T.) KHABAZA, (T.) REINARTZ, (C.) SHEARER, (R.) WIRTH - CRISP-DM 1.0 step-by-step data mining guide, CRISP-DM Consortium, August 2000.
- (2) (T.) KOHONEN, (J.) HYNINEN, (J.) KANGAS, (J.) LAAKSONEN - SOM_PAK, The Self-Organizing Map program package, technical report A31, Helsinki University of Technology, 1996. <http://www.cis.hut.fi/nnrc/nnrc-programs.html>, 1996.
- (3) (J. -H.) FRIEDMAN - Multivariate adaptive regression splines - *The Annals of Statistics*, vol. 19, n°1, 1-141, 1991.
- (4) (M. -T.) RODRÍGUEZ MONTEQUÍN - Modelado evolutivo mediante técnicas adaptativas aplicado al control de inclusiones en bobinas laminadas en caliente, PhD thesis, 2003.
- (5) (M. -T.) RODRÍGUEZ, (F.) ORTEGA, (J.-L.) RENDUELES, (C.) MENÉNDEZ - Combination of multivariate adaptive techniques and neural networks for prediction and control of internal cleanliness in steel strips, proceedings of EUNITE 2003.