ISSN 1983-4764

DEFECT MODELING: AN ADVANCED APPROACH TO SOLVE QUALITY PROBLEMS[1]

Fabricio Brandão Pereira de Souza[2]
Tiago da Silva Fernandes[3]
Luciano Colzani[4]
Danilo Gonçalves de Oliveira[5]
Ram Patil[6]

Abstract

Many industrial plants are still using simple and popular tools such as spreadsheet applications to analyze complex and multi-variable problems, which do not take into account the probabilistic relations. Excessive time consumption to solve problems, productivity loss, unreliable results and dimensional constraints can be stated as typical limitations of these tools. To eliminate these drawbacks, ArcelorMittal has successfully developed a new methodology that extracts high level information within production database and uses unconventional statistical tools based on Bayesian Network (BN) Modeling for addressing production issues. ArcelorMittal's VEGA facility has deployed internally this new approach to perform defect analysis in a more efficient and effective manner.

Key words: Defect modelling; Bayesian network; Quality defect; Galvanizing line.

[1] *Technical contribution to the 49th Rolling Seminar – Processes, Rolled and Coated Products, October, 22nd-25th, 2012, Vila Velha, ES, Brazil.*
[2] *Mechanical Engineer, Hot Dip Galvanizing Process Specialist - ArcelorMittal VEGA. São Francisco do Sul – Santa Catarina – Brazil.*
[3] *Hot Dip Galvanizing Process Specialist – ArcelorMittal VEGA. São Francisco do Sul – Santa Catarina – Brazil.*
[4] *Pickling and Cold Mill Area Manager – ArcelorMittal VEGA. São Francisco do Sul – Santa Catarina – Brazil.*
[5] *Electrical Engineer, Cold Mill Process Specialist – ArcelorMittal VEGA. São Francisco do Sul – Santa Catarina – Brazil.*
[6] *Coating and Finishing Process Portfolio - ArcelorMittal Global R&D. East Chicago – Indiana –USA.*

ISSN 1983-4764

1 INTRODUCTION

Many industrial plants are still using simple and popular tools such as spreadsheet applications to analyze complex and multi-variable problems. Excessive time consumption to solve problems, productivity loss, unreliable results and dimensional constraints can be stated as typical limitations of these tools. ArcelorMittal VEGA has deployed internally a new approach to perform defect analysis in order to eliminate the drawbacks stated above.
Advantages of using data mining tools instead of simple analysis tools currently used on industrial plants such as Excel and others are:

- Excel is not really appropriate because it can only do simple analysis, plot variables in one or two dimensions and make linear regression and not much else. It is very difficult in Excel to identify correlation between process variables and defect measurement. The main reason for this is that defects have, in general, non-linear behaviour in dimension greater than two.
- Moreover, it is better to use probabilistic models because generally, not all the relevant information is available in a database. Nevertheless, one can calculate the probability of occurrence of a defect for a given point. Excel is not able to take into account this probabilistic behaviour.
- Time saving: Requires only some minutes to extract the main patterns instead of hours or days.

The methodology for conducting a defect crisis analysis is founded on the following pillars (Figure 1): problem identification; defect characterization; process and product data collection; modeling (build Bayesian network model); Model evaluation and hypotheses testing – and then industrial trials; and finally, solution deployment and process control. Data analyses are performed using the commercial software Bayesialab – data mining software based on Bayesian Network modeling.

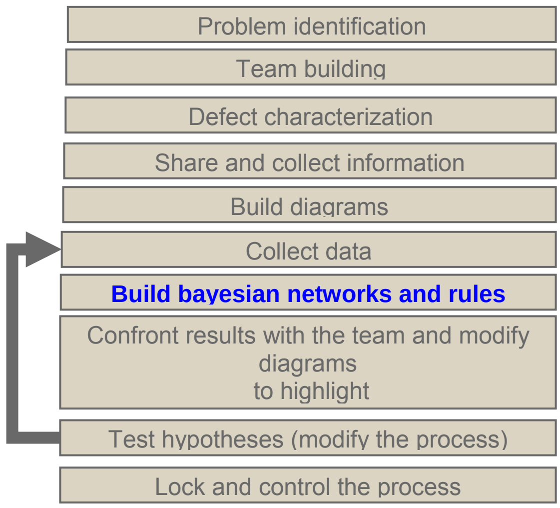

Figure 1. Steps for conducting a defect crisis analysis.

The major key of analysis success relies on the database building. A short description of the main steps to collect and build the database is presented below:

- Collect process data: that has been highlighted in the previous analysis
  - Identify a stable period for data sampling
  - Make sure the problem arises often enough

ISSN 1983-4764

- Select a measurement frequency in accordance with the failure mode of the response
- Collect all parameters listed during the process mapping

- Deal progressively with local information (from the process to upstream lines information).
- Analyze data and clean outliers if they really are.
  - Visualize each variable distribution and look at extreme values
  - Identify cases with lowest probability value with the Bayesian network (this is done in the further steps).
- A prerequisite to find influent variables is to have data with defect and data without defect (if not, a variation in defect density should be observed in the data).

A BN is a graphic probabilistic model through which one can acquire, capitalize on and exploit knowledge. It combines the rigor of powerful and stable mathematical formalism, the effectiveness of a distributed representation of knowledge and the readability of rule-based models. It is used to represent knowledge from a system or to find out this knowledge by analyzing data (learning). Through the network one can diagnose and simulate, analyze data, make decisions and control systems.

2 BAYESIAN NETWORK

A Bayesian network, belief network or directed acyclic graphical model is a probabilistic graphical model that represents a set of random variables and their conditional dependencies via a directed acyclic graph (DAG). Probabilistic graphical models are graphs in which nodes represent random variables, and the (lack of) arcs represent conditional independence assumptions. Hence they provide a compact representation of joint probability distributions.
Figure 2 illustrates a simple BN. Edges represent conditional dependencies; nodes which are not connected represent variables which are conditionally independent of each other. Each node is associated with a probability function that takes as input a particular set of values for the node's parent variables and gives the probability of the variable represented by the node.

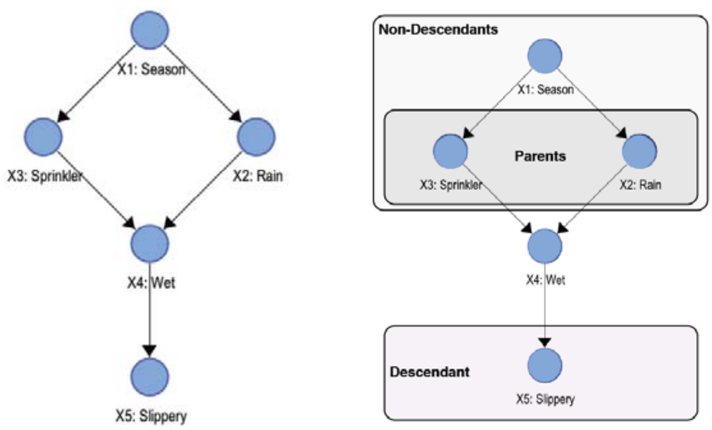

Figure 2. Direct Acyclic Graph – Bayesian Network Illustration.

Any complete probabilistic model of a domain must, either explicitly or implicitly, represent the joint probability distribution— the probability of every possible event as

ISSN 1983-4764

defined by the combination of the values of all the variables according to the chain rule:[(1)]

$$P(x_1,\ldots,x_n)=P(x_1)\prod_{i=2}^{n}P(x_i|x_1,\ldots,x_{i-1}) \quad (1)$$

There are exponentially many such events, yet Bayesian networks achieve compactness by factoring the joint distribution into local, conditional distributions for each variable given its parents. If x_i denotes some value of the variable X_i and pa_i denotes some set of values for the parents of X_i, then $P(x_i|pa_i)$ denotes this conditional distribution.

$$P(x_i,\ldots,x_n)=\prod_{i}P(x_i|pa_i) \quad (2)$$

In the network of figure 2, we have

$$P(x_1,x_2,x_3,x_4,x_5)=P(x_1)P(x_2|x_1)P(x_3|x_1)P(x_4|x_2,x_3)P(x_5|x_4) \quad (3)$$

2.1 Difference Between Prediction and Control

We remind some points that appear to be important in this problematic, the difference between prediction and control, and illustrate it with a simple example.
In prediction, the interpretation is based with no change in the observed universe. A predictor is constructed with a set of observations and prediction on a new observation is calculated without changing the process that has generated the previous set of data. In this context, the knowledge of the probability of the observations is sufficient to infer a prediction.
In control, the situation is more difficult as we try to find some actions on the system that will attain a particular objective. In this context, we are facing the problem of constructing a causal model from data. If the real cause is not measured in the database, it is even possible to obtain rules that seem to have good efficiency. To highlight the difference, we present a simple example with a model containing three variables:

- C has two states C1 & C2, which is the parameter responsible for the defect that cannot be-observed
- P has two states P1 & P2, an observed parameter, which is dependant on the parameter C
- D representing the defect with Defect or Ok states. The marginal probability of the state Defect is 12.4% and Ok is 87.6%.

We use a Bayesian network notation to represent this model:

ISSN 1983-4764

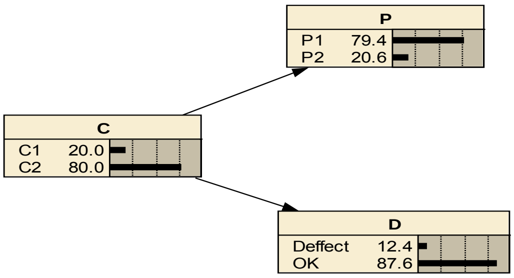

Figure 3. Distribution for each parameter in the database (ex : C is 20% at C1).

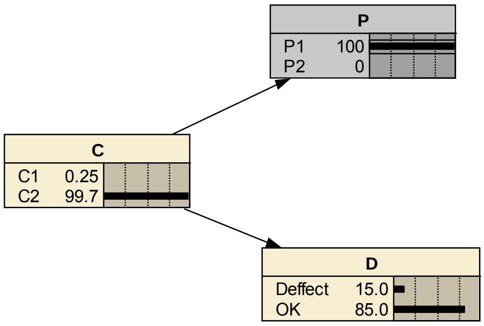

Figure 4. In the database with P = P1, the rate of defect is 15%.

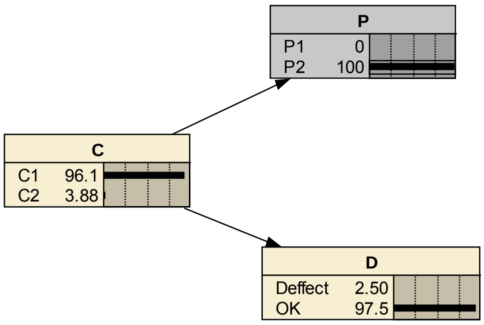

Figure 5. In the database with P = P2, the rate of defects is 2.50%.

In prediction mode, if we observe P=P1 (Figure 4), the risk of defect is 15% and, otherwise, if P = P2 (Figure 5) the risk of defect is 2.5%.
With a database obtained from this model, a rule-based model will produce the rule: if P = P2 then purity is 97.5% with a coverage of 20.6%.
In a control mode, if we impose on P to be at the value P2, it is not at all guaranteed that the rate will decrease at a rate of 2.5%. Indeed, in this model it is C which has a physical effect on P and not the opposite.
This little example shows that it is not possible to guarantee with a rule-based algorithm causal relationship between process parameters and defect variable. In real case, the cause is the consequence of many conjugate effects and this problem is increased by the number of variables to take into account.
To our knowledge, a way to infer causal relationship is to adopt for instance the theory proposed by Pearl.[2] In this theory, directed graphs (Bayesian Network) are constructed with some algorithms based on conditional independence between variables. In this context, it is then possible to propose a family of possible graphs with causal relations.

ISSN 1983-4764

2 EXAMPLE OF DEFECT ANALYSIS – EDGE WAVES ON HDG PRODUCTS

The Edge Waviness Defect is the result of an abnormal elongation at the edges of the strip. The effects of the defect is the downgrading of the material or reallocation to another customer requesting a narrower strip, involving non-scheduled extra processing such as edge cutting the material, incurring additional yield losses and increased cycle time in supplying the original customer.
The first root cause assumptions were related to the flatness of the full hard coils and/or zinc over coating at the edges of the strip at the continuous galvanizing line. On the defect analysis, many hours and energy have been spent by process and production teams to identify the root causes, parameters to change and control, but without success. Hence, the defect modeling approach was undertaken.
In order to clarify the results described in the next section, a short description of the tandem cold mill is appropriate: The Tandem Cold Mill (TCM) at ArcelorMittal Vega is a four stands, 4High mill equipped with Bending and CVC (Continuously Variable Crown) as flatness controllers. The CVC consists of hydraulic equipment moving axially both top and bottom work rolls, which are ground to an S-Shaped curve. As a result, the strip is subject to work rolls with crown varying from 0 to 0.45µm continuously, requiring less work roll inventory and easier roll-shop management. On the other hand, as it is a new technology and not very common around the world, it is still under development to determine good presets for Bending and CVC to improve flatness of full hard coils going to the galvanizing line.

3 ANALYSIS WITH DEFECT MODELING APPROACH

One very useful property of BN is its ability to learn from observations. Learning of BN can be divided into two types: parameter learning and structure learning. With parameter learning, the structure of the BN is given and only the conditional probability table parameters are learned. With structure learning, the BN structure itself is learned.
Structure Learning: Automatic algorithms allow the expert to quickly discover all the direct probabilistic relationship in the data with a global view, thanks to the graphical representation of the Bayesian Network model. Parameters are positioned according to the probabilistic relationship force (Figure 6). Network structure learnt is in accordance with process flow as well as parameters relationship as highlighted through the different inner nodes colors (clustering technique) – HDG parameters located in the red bound box and TCM parameters in the blue bound box.

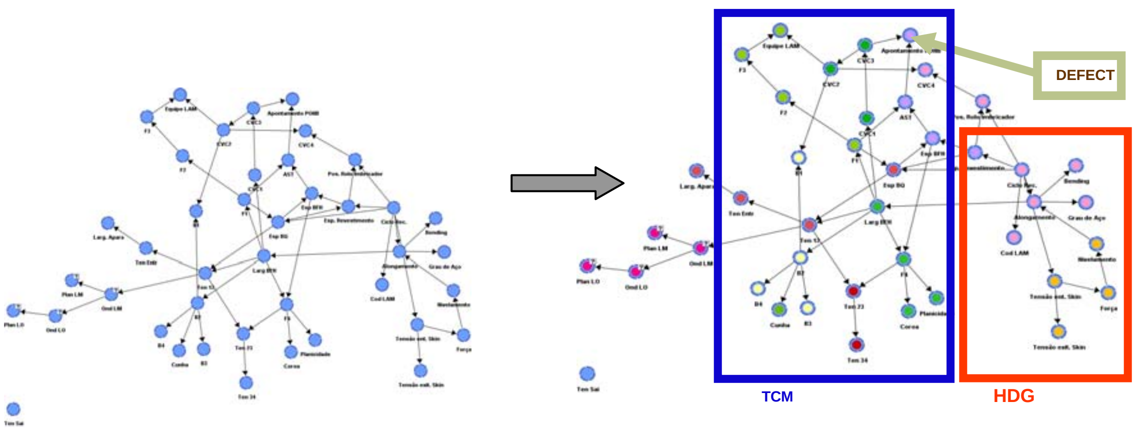

Figure 6. Bayesian Network - Discover the direct probabilistic relations in the data.

ISSN 1983-4764

Defect parameter is located in the surrounding area of TCM variables, giving to the expert a first overview of the overall relationship.
Supervised learning for defect modeling: one can focus entirely on the characterization of the target variable, e.g. defect occurrences, aiming to identify the sufficient set of variables that can explain the defect and the probabilistic relations (Markov blanket). The Markov blanket has a very interesting property. It is made up of the parents, the children and the co-parents. If the values of the nodes are known, the target node becomes entirely independent of all the other variables.
Additionally, the model created can be useful for automatic defect classification and support on-line decision making.
The main features stressed by the network (Figure 7) concern cold rolling parameters, mainly those controlling the crown of the work rolls, related to the flatness control. Markov blanket is an interesting tool for extracting the most important variables (feature selection).

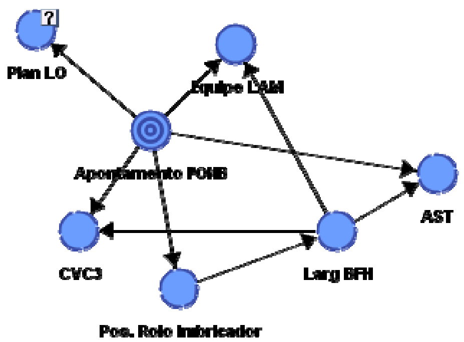

Figure 7. Bayesian Network – Defect characterization.

Probabilities distributions of each parameter are shown in the Figure 8.

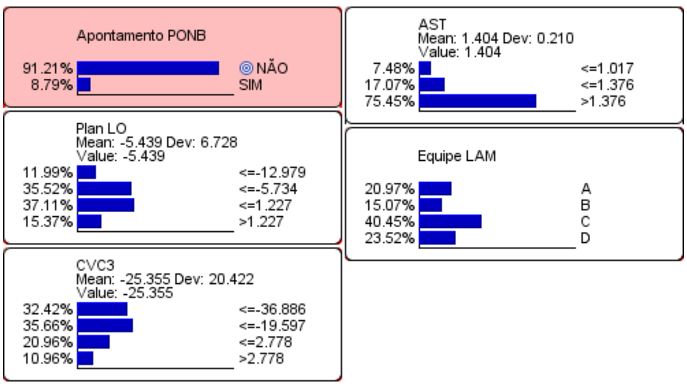

Figure 8. Probability distributions are clearly visualized for a better understanding.

4 INFERENCE AND PROBLEM SOLVING

Because a Bayesian network is a complete model for the variables and their relationships, it can be used to answer probabilistic queries about them. For example, the network can be used to find out updated knowledge of the state of a subset of variables when other variables (the evidence variables) are observed.
Dynamic inference can thus be performed. Inference is the process of calculating the probability of one or more variables “X” given some evidence “e”. The evidence is

ISSN 1983-4764

expressed as an observation of the state of some variables in the BN. In short: *P(X|e)* needs to be calculated.(3)
Because of the propagation of the new information (defect=yes) the probability distributions for every node have been updated, allowing the expert to evaluate the new probability distributions (Figure 9).

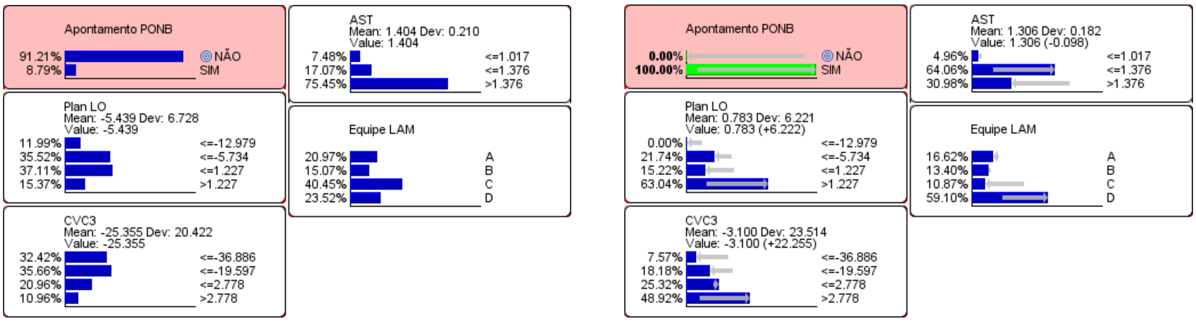

Figure 9. Dynamic inference – evidence is observed and probabilities are updated.

The conditional probability table of the defect variable shows that 59.10% of the coils with defect were produced by the operational team "D" and 48.92% of coils were produced with the CVC parameter of the stand number three above 2.778 mm. In consequence of that practice, flatness issues have been highlighted through the parameter "Plan LO".
Discrepant results have been observed when comparing cold rolling parameters (mainly bending and CVC) and production team. A large amount of coils produced by a specific shift that are directly related to the parameters above can explain the defect occurrences, e.g. differences in operating practices among the operators. Corrective actions have been taken based on these results, actuating on the CVC position and bending force parameters aiming to get a more appropriate full hard flatness for further galvanizing process.
The first analysis carried out based on the data collected during the defect occurrences increase in May/10 have shown an expressive influence of the CVC position of Stand 3 at the TCM. After actions taken, material loss dropped from 0,54% to 0,19% in the next two months.
Because of the complexity of controlling flatness through a four-stand cold rolling mill with both CVC and bending control, any process perturbation at the TCM or upstream lines is sufficient to increase defect occurrences. This could be proven four months after the first set of actions taken. A new data base and analysis have established presets for Bending and CVC of Stands 1, 2 and 3 of TCM. Moreover, an accurate prediction model was obtained, classifying correctly up to 85% of defective coils (using a test database). Material loss rate dropped from 1,17% to less than 0,15% after second set of actions taken, as can be seen on Figure 10.

ISSN 1983-4764

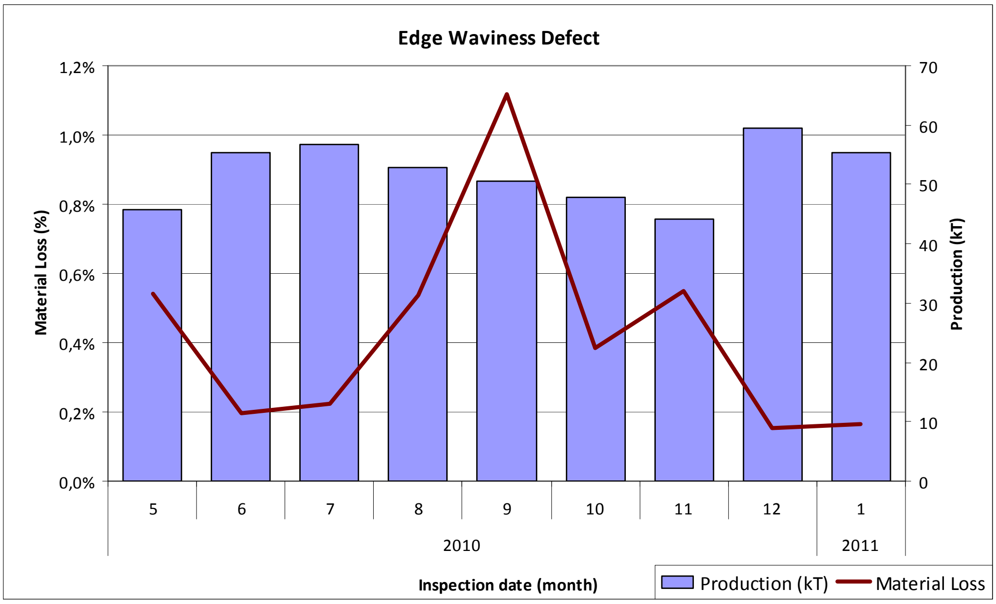

Figure 10. Material loss and HDG production.

5 CONCLUSION

A methodology has been successfully developed to extract high level information within production database in order to solve production issues, such as quality problems and support decision making by using unconventional statistical tools based on Bayesian Network Modeling.
To recap, after an accurate database is built, a detailed analysis can be carried out in a few minutes by using proper statistical methods, saving time and allowing the experts to diagnose and simulate the system behaviour and support decision making.
Some advantages of using BN formalism for these purposes are:

- Allow one to learn about relationships between variables and as it is graphical it facilitates the understanding.
- Is robust as it is statistically generated and the complexity can be controlled.
- BN is able to model causal relationships between variables.
- Can handle incomplete data sets and it is possible to combine expert knowledge and data into a BN.
- If needed the model can be integrated in a production line.

Besides defect diagnosis, internal studies are being carried out in order to master strip flatness control and process flexibility.

REFERENCES

1 BARBER, D., Bayesian Reasoning and Machine Learning. Cambridge University Press, UK, 2011. chapter 3 , p. 31 – 55.
2 PEARL, J., Causality: Model Reasoning and Inference, Cambridge University Press, New York, 2000. 400 p.
3 HULST, J., Modeling Physiological Processes with Dynamic Bayesian Networks, PhD Thesis, Delft University of Technology, Netherland, 2006. chapter 3, p. 15 – 33.