Monitoring of Casting Quality Using Principal Component Analysis Based Self-Organizing Map

Hocine Bendjama (✉ h.bendjama@crti.dz)

Research Center in Industrial Technologies CRTI

Salah BOUHOUCHE

Research Center in Industrial Technologies CRTI

Salim AOUABDI

Research Center in Industrial Technologies CRTI

Jürgen BAST

IMKF, TU Bergakademie Freiberg

Research Article

Keywords: Condition monitoring, Principal component analysis, Self-organizing map, Casting quality, Hotelling's T², Q statistic

Posted Date: November 9th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1036731/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Version of Record: A version of this preprint was published at The International Journal of Advanced Manufacturing Technology on March 7th, 2022. See the published version at <https://doi.org/10.1007/s00170-022-08993-9>.

Monitoring of casting quality using principal component analysis based self-organizing map

Hocine BENDJAMA¹, Salah BOUHOUCHE¹, Salim AOUABDI¹, Jürgen BAST²

¹Research Center in Industrial Technologies CRTI, P.O.Box 114 Cheraga Algiers Algeria

²IMKF, TU Bergakademie Freiberg, Cotta Strasse 4, D-9596, Germany.

Corresponding author: h.bendjama@crti.dz

Abstract

The monitoring of casting quality is very important to ensure the safe operation of casting processes. In this paper, in order to improve the accurate detection of casting defects, a combined method based on Principal Component Analysis (PCA) and Self-Organizing Map (SOM) is presented. The proposed method reduces the dimensionality of the original data by the projection of the data onto a smaller subspace through PCA. It uses Hotelling's T^2 and Q statistics as essential features for characterizing the process functionality. The SOM is used to improve the separation between casting defects. It computes the metric distances based similarity, using the T^2 and Q (T^2Q) statistics as input. A comparative study between conventional SOM, SOM with reduced data and SOM with selected features is examined. The proposed method is used to identify the running conditions of the low pressure lost foam casting process. The monitoring results indicate that the SOM based on T^2Q as feature vectors remains important comparatively to conventional SOM and SOM based on reduced data.

Keywords: Condition monitoring; Principal component analysis; Self-organizing map; Casting quality; Hotelling's T^2 ; Q statistic

1. Introduction

Nowadays, the monitoring of casting quality plays a very important role to ensure the steadiness of casting processes. Its main objective is to ensure the dependability of processes and increase their availability with lower cost. Casting technology is one of the oldest known metal-forming techniques in which the molten metal is poured into a mold where it cools and solidifies. Due to its wide use in the foundry industry, modern developments have led to a broad array of specialized casting processes to create complex castings at a high production rate and good quality. Even in these controlled processes, defects in the output products can occur.

Generally, casting defects fell into one or more of the established seven categories of casting defects such as: metallic projections, cavities, discontinuities, defective surface, incomplete casting and incorrect dimensions or shape [1]. The formation of casting defects is linked to several factors like pouring temperature, metal velocity, sand and refractory coating [1, 2]. Obviously, most castings suffer rejection due to defect formation. It is therefore essential or necessary to detect the casting defects in order to improve the performance of process monitoring. Traditionally, non-destructive testing techniques (eddy current, magnetic leakage and so on) have been used and applied with human participation [3, 4]. However, these manual techniques are usually exhausting, error-prone, and expensive. Today, many computer system-based defect detection methods bring great convenience for quality control and provide operators with the process operating conditions.

Process condition monitoring is not an easy task; it is essentially a problem of process knowledge. In current methods, after data collection, features are first extracted and generated, which often require human expertise. Then the process identification takes place. The most effective feature extraction and more accurate identification are needed to obtain the correct operating condition of the process.

In recent years, the use of combined and hybrid forms between different monitoring methods is an effective tool to improve the identification accuracy. It requires a critical processing step designed to find the most informative feature. Various combined methods have been widely developed and applied to identify casting defects, e.g. [5] presented a new combination method of casting defect analysis based on Taguchi method and computer aided casting simulation technique to analyse the sand related and methoding related defects in green sand casting. [6] optimized the low pressure die-cast process by a combining neural network and genetic algorithm method. [7] extracted a robust randomly distributed triangle feature for casting defects detection such as; cracks, blow holes, shrinkage porosities and shrinkage cavities. [8] developed a casting defect analysis system to analyse and reduce defects in investment castings based on Bayesian inference with pre-processing of the input data. [9] proposed an inspection system based on computations of independent quantities on several areas within an image to identify defects on the surface of a metal part produced by a casting process. [10] developed a fault detection module based on an artificial neural network with the pre-processing of time series temperature and pressure measurement data to reduce the defect rate and production cost in die-casting industry. [11] embedded a deep learning algorithm in a digital camera to quickly and automatically detect casting defects. To effectively detect surface defects in continuous casting process, a methodology for automatic feature extraction and classification using vision-based sensing technology have been presented by [12]. Similarly, [13] proposed a robust detection

method of casting defects based on vision attention mechanism and deep learning of feature map. [14] used contourlet transform and kernel locality preserving projections to extract sufficient and effective features from metal surface data. [15] developed a predictive control strategy using a heat transfer model and Quasi-Newton method to improve product quality in continuous casting. In another application, [16, 17] extracted a reduced data matrix by Principal Component Analysis (PCA) method, which is used as input to Self-Organizing Map (SOM) algorithm to evaluate the pickling process. [18] suggested a combined simulation approach between Monte Carlo method and neural networks to bring essential information on the state of materials. The work presented in this paper is motivated by practical applications in the foundry industry, where a hybrid form of PCA-SOM is used to defect assessment.

The multivariate statistical method PCA has become widely used to improve fault detection accuracy in recent years. PCA is a data compression method; it produces a lower dimensional representation in a way that preserves the correlation structure among the process variables. The collected casting measures are used as inputs to the PCA algorithm, which is in charge of the detection of abnormal situations and provide information about the state of the process using Hotelling's T^2 and Q statistics. These statistical parameters must do not exceed threshold values in normal situation, however, sometimes T^2 and Q statistics alone could not detect the faulty conditions, so additional algorithms are necessary.

The proposed hybrid form uses the advantages of the two proposed methods; PCA and SOM. PCA is used for data pre-processing, data reduction and feature extraction. The features are obtained from casting data analysis using T^2 and Q (T^2Q) statistics, which represent the input vectors to the SOM algorithm. The SOM computes the metric distances based similarity in order to identify the running conditions of the low pressure lost foam casting process. The main goal of this proposed method is to obtain more detailed information contained in the measured data than had been previously possible. The monitoring results using real measurement data demonstrate that casting defects can be clearly identified by utilizing the proposed method.

2. Considered method

In order to construct a successful casting defects identification system, a combination using PCA and SOM, is described in this paper (see Fig. 1). PCA is trained with input matrix that contains the process variables, where the goal is to establish the normal statistical correlation among the measured data to characterize the operating conditions of the casting process using Hotelling's T^2 and Q statistics. Sometimes, these statistics cannot exactly detect process faults, they are proposed here as characteristic features of casting data. The chosen features are used as input to SOM algorithm, which is used for casting quality evaluation. The main objective is to compute the distances between the input vectors and winning neuron to better identify casting defects. A comparative study between conventional SOM, SOM with Reduced Data matrix (RD) and SOM with selected features (T^2Q) is examined.

Fig. 1 Diagram of the proposed method

2.1. Principal Component Analysis

The PCA [19] is a multivariate statistical analysis technique, which reduces the original data space into a smaller-dimension space in terms of protecting the main original data information. Given data matrix $X \in \mathcal{R}^{m \times n}$ composed of m observations or samples and n variables, which have been normalized to have zero mean and unit variance, PCA is only interested in its variance and covariance. PCA actually relies on eigenvalue/eigenvector decomposition of the covariance or correlation matrix C given by:

$$C = (1/(m-1))X^T X = PDP^T \quad (1)$$

where $D = \text{diag}(\lambda_1, \dots, \lambda_n)$ is a diagonal matrix with diagonal elements in decreasing magnitude order and P contains the eigenvectors.

PCA determines an optimal linear transformation of the data matrix X in terms of capturing the variation in the data as follows:

$$T = XP \quad \text{and} \quad \hat{X} = TP^T \quad (2)$$

Where $T \in \mathcal{R}^{m \times k}$ is the principal component matrix and the matrix, $P \in \mathcal{R}^{n \times k}$ contains the principal vectors which are the eigenvectors associated to the eigenvalues λ_i of the covariance matrix and k denotes the Principal Components number (PCs) of PCA model. A key issue to develop a PCA model is to choose the adequate number of PCs. A number of well-known techniques have been proposed for selecting the number of PCs [20]. In this study, the specific calculation method of selected k follows the Cumulative Percent Variance (CPV) principal [21]. It is a measure of the percent variance, such as 85%, captured by the first PCs.

The difference between X and $\hat{X}$ is the residual matrix E . It can be calculated as:

$$E = X - \hat{X} = X(I - PP^T) \quad (3)$$

Where I is the unit matrix.

To perform process fault detection, a PCA model of the normal operating conditions must be built. When a new observation data is subject to faults, these new data can be compared to the PCA model. The correlation of the new data is detected by Hotelling's T^2 and Q , called also Squared Prediction Error (SPE), statistics as follows:

$$T^2 = X^T P D^{-1} P^T X \quad (4)$$

$$Q = E^T E = (X - \hat{X})^T (X - \hat{X}) \quad (5)$$

The process is considered normal if Hotelling's T^2 and Q statistics do not exceed threshold values. These statistics alone could not detect the faulty conditions. They are used, in this study, as input to the SOM algorithm to further improve the separation between casting conditions.

The computing steps using PCA method are summarized as follows:

- (1) Matrix X ,
- (2) Mean centring of X ,
- (3) Calculation of covariance matrix,
- (4) Eigenvalues and eigenvectors,
- (5) Optimal number of PCs,
- (6) PCA Model,
- (7) T^2 and Q statistics.

2.2. Self-Organizing Map

The SOM, invented by Kohonen [22], is a kind of artificial neural network that uses unsupervised competitive learning to map a high dimensional input space (the data space) onto a low dimensional output space, usually of two dimensions. It is a technique to group data with similar characteristics. Each neuron or node comprises a vector of weights of the same dimension as the input data vectors. The SOM is trained by presenting the data repeatedly and upgrading the weights to learn the structure of the data.

The SOM consists of an input layer and a competitive or output layer, fully interconnected to each other. The output layer consists of m neurons. Each neuron i ($i = 1, 2, \dots, m$) is represented by an n dimensional weight vector $w_i = [w_{i1}, \dots, w_{in}]$ where n is the dimension of the input vector. In the output layer, the competitive process is done and the weight of connection is updated to choose a winner neuron.

The key steps in the SOM learning process (training) are first, for each input vector, determining its Best Matching Unit (BMU). The BMU is the node that is most similar to the input vector. If we denote b the BMU of input vector x and w_b the weight vector of this BMU, the identification is based on the minimum Euclidian distance, which is defined as:

$$\|x - w_b\| = \min_{i=1, \dots, m} \{\|x - w_i\|\} \quad (6)$$

Then the process proceeds to update the weight vectors of the BMU and those of its adjacent neurons (neighborhood) to match the input data using the following equation:

$$w_i(t+1) = w_i(t) + \alpha(t) h_b(x)_i(t) [x(t) - w_i(t)] \quad (7)$$

Where t denotes time, $0 < \alpha(t) < 1$ is the learning rate, $h_b(x)_i(t)$ is the neighborhood function centered in the BMU b at time t . The neighborhood function is often taken as a Gaussian function. Both $\alpha(t)$ and the width of $h_b(x)_i(t)$ decrease gradually with the increasing step t .

After the training step, the winner neuron is used to compute metric distances. Distances or similarity measures are essential to solve many pattern recognition problems such as classification and clustering. Various distance measures that are applicable. In this work, we consider a similarity evaluation based on the Euclidian distances [16, 17]. There will be used different formula to measure distances between two data points characterizing two monitoring conditions. The followings distances can be applied:

- Distance between input and neuron data:

$$DIN(x, w_{bj}) = \sqrt{\sum_{j=1}^n (x_j - w_{bj})^2} \quad (8)$$

- Mean or gravity center of the input data:

$$GCI = (1/n) \sum_{j=1}^n x_j \quad (9)$$

- Gravity center of the neuron data:

$$GCN = \frac{\sum_{j=1}^n w_{bj} x_j}{\sum_{j=1}^n w_{bj}} \quad (10)$$

- Distance between gravity centers (input and neuron data):

$$OGX = \sqrt{\sum_{j=1}^n (GCI - GCN)^2} \quad (11)$$

Where x_j is the input vector and w_{bj} defines the winner neuron.

The decision about running conditions can be performed using the computed metric distances; DIN and OGX, separating the current operation data and process conditions. The differences between two running conditions i and j ($\Delta DIN = DIN_i - DIN_j$ and $\Delta OGX = OGX_i - OGX_j$) are used as indexes to distinguish the defective and healthy casting conditions. The SOM algorithm is described as follows:

- (1) Set the initial random positions and the initial learning rate,
- (2) Provide an input data,
- (3) Computing and finding BMU,
- (4) Adjust the weights,
- (5) Construct the output space,
- (6) Compute similarity measures; DIN and OGX.

3. Experiment setup and data acquisition

The measurement data applied to condition monitoring requires different types and levels of equipment and techniques. These depend on the investment and available expertise.

The experimental measurements presented in this paper are entirely based on the casting data acquired from the low pressure lost foam casting process. As shown in Figure 2, the casting process utilizes a resistance furnace capable of melting standard Aluminium base alloys. Air pressure is applied to the chamber containing the crucible to push liquid metal up into a flask containing the foam pattern and unbounded sand. More details about the casting process were reported in [23, 24].

The low pressure lost foam casting process is utilized to create complex castings. Filling the mold drives the molten metal into the pores of the sand surface and produces the undesired mold casting. Generally, a casting defect is defined as all observable and unplanned variation. If a defect occurs, measures must be adopted to control the quality of castings. The data used in this study were gained through the experimental measurements in normal (NR) or healthy operating conditions and abnormal functionality of the casting process including three defect; cracks (SC), holes (SH) and metal penetration (MP).

Fig. 2 Schematic representation of the low pressure lost foam casting process

To acquire data, five temperature transducers were implemented in the process by five thermocouples for temperature input. The pressure and temperature inputs were wired to a National Instruments (NI) data acquisition board with a USB interface. The foam pattern was supported in the flask and the thermocouples were wired to the data acquisition unit. When the vessel is pressurized, liquid metal rises through a steel pipe into the

flask. All test part was cast using AlSi12 alloy at temperatures between 725°C and 750°C (see Table 1). NI DASY Lab software was used to collect and analyse the signals from the temperature sensors. The measured variables are presented in Fig. 3.

Table 1 Casting conditions

Test	Pouring Temperature (°C)	Measured Temperature (°C)	Holding Pressure (bar)	Mold Filling time (s)	Holding time (s)
NR	735	711	0.24	6	90
SC	750	700	0.24	6	90
SH	730	730	0.24	6	90
MP	725	725	0.24	6	90

Fig. 3 Measured data

4. Results and discussion

The structure of the proposed fault identification method involves two parts; one is the development and training the models, the other is testing the process fault based on trained models. The measured data used in training represent a healthy casting condition.

In this study, the data matrices that contain the values of measured variables are constructed from 930 sampling instances of interval 0.1s. They represent the experimental measurements acquired from normal (NR) and faulty (SC, SH, MP) process operations. In total, 8 variables and 930 data points at different times are collected for training the different models. The obtained matrices are used to evaluate the performance of the SOM and PCA-SOM techniques, the corresponding indexes are thus calculated and discussed.

Through PCA, the eigenvalues of the covariance matrix, which are the variances of PCs, are listed in Table 2. The anterior 2 PCs explain over 85 of the total variance of the data. The PCA model is established making use of them, and then the monitoring performance is progressed. As shown in Figure 4, all process variables are correctly estimated with this PCA model, except that some certain variables are less well estimated than others.

Table 2 Eigenvalues and variances of PCs

PCs	Eigenvalues	Variances (%)
1	6.4363	80.45
2	1.0164	12.71
3	0.2960	3.70
4	0.1143	1.43
5	0.0700	0.87
6	0.0395	0.50
7	0.0199	0.24
8	0.0076	0.10

Fig. 4 Estimation of the measured data

According to SOM algorithm and PCA-SOM with RD as input (PCA-SOM/RD), casting conditions are evaluated by the metric distances DIN and OGX given by Eqs. (8) and (11) respectively. Obtained results are represented in Figs. 5 and 6. As we can see, it is not easy to distinguish between normal and abnormal conditions and also between false and missed alarms.

Fig. 5 Evolution of DIN and OGX indexes using SOM algorithm

Fig. 6 Evolution of DIN and OGX indexes using PCA-SOM/RD technique

The proposed algorithm, PCA-SOM with T^2Q as input (PCA-SOM/ T^2Q), is tested using the same computed indexes. The isolation level of the normal and abnormal casting conditions is clearer than the SOM and PCA-SOM/RD. The tested distances indicate that the casting defects are successfully identified (see Fig. 7). The output results can be used to describe the capability of fault isolation.

Fig. 7 Evolution of DIN and OGX indexes using PCA-SOM/ T^2Q technique

To evaluate the performance of the SOM, PCA-SOM/RD and PCA-SOM/ T^2Q techniques, a comparative study using ΔDIN and ΔOGX distances is carried out. Differences between computed indexes of defective and healthy casting conditions are listed in Table 3. From obtained results, important variations are detected using PCA-SOM/ T^2Q . These variations show the detection capacity of casting conditions. Moreover, SOM and PCA-SOM/RD techniques take more computing time than PCA-SOM/ T^2Q . The proposed method therefore shows a significant high performance compared to the considered methods (SOM and PCA-SOM/RD).

Table 3 Performance evaluation

Defects Methods	Δ DIN			Δ OGX			Computing time (s)
	SC	SH	MP	SC	SH	MP	
SOM	27.45	-	-	-	-	-	54.33
PCA-SOM/RD	7.06	1.16	3.81	0.57	1.89	24.58	55.62
PCA-SOM/T ² Q	148.74	50.38	97.16	84.91	25.32	42.69	50.80

On the other hand, the proposed method also proves its ability to ensure high defect detection performance compared with other applications that use PCA-SOM with RD as feature [16, 17].

5. Conclusion

In this paper, a new statistical method to improve the monitoring of casting conditions based on a combination of PCA and SOM is proposed. The PCA is exploited to extract statistical features from experimental data. Using Hotelling's T^2 and Q statistics as input to SOM algorithm, the casting defects are separated using Euclidian distances based similarity evaluation. This combined method is accurate in casting quality monitoring, it can be seen from the analysing results that the normal and the abnormal casting conditions can be clearly identified. It is clear that the proposed method outperforms the other approaches and algorithms considered in this work.

Data availability

Not applicable.

Code availability

Not applicable.

Acknowledgments

This research acknowledges the German Academic Exchange Service (*in German*. Deutscher Akademischer Austausch Dienst (DAAD)). The authors would like to extend their greatest gratitude to the Institute of Mechanical Engineering, Freiberg University of Mining and Technology, Freiberg, Germany. Comments and suggestions from anonymous reviewers and the Editor are greatly appreciated to improve the manuscript.

Author information

Affiliations

1. Research Center in Industrial Technologies CRTI, P.O.Box 114 Cheraga Algiers Algeria.

Hocine BENDJAMA, Salah BOUHOUCHE, Salim AOUABDI

2. IMKF, TU Bergakademie Freiberg, Cotta Strasse 4, D-9596, Germany.

Jürgen BAST

Corresponding author

Correspondence to Hocine BENDJAMA.

Ethics declarations

Ethics approval

Not applicable.

Consent to participate

Not applicable.

Consent for publication

Not applicable.

Conflict of interest

The authors declare no competing interests.

Additional information

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Beeley P (2001) Foundry Technology. 2nd ed. Oxford: Butterworth-Heinemann
2. Rao P (2008) Manufacturing Technology: Foundry, Forming and Welding. Tata McGraw Hill
3. Hughes RR, Dixon S (2014) Eddy-current crack detection at frequencies approaching electrical resonance. In: AIP Conference Proceedings 1581 (1): 1366–1373
4. Kim HM, Rho YW, Yoo HR, Cho SH, Kim DK, Koo SJ, Park GS (2012) A study on the measurement of axial cracks in the magnetic flux leakage NDT system. In: IEEE International Conference on Automation Science and Engineering (CASE): 624–629
5. Dabade UA, Bhedasgaonkar RC (2013) Casting Defect Analysis using Design of Experiments (DoE) and Computer Aided Casting Simulation Technique. Procedia CIRP (Forty Sixth CIRP Conference on Manufacturing Systems) 7: 616–21

6. Zhang L, Wang R (2013) An intelligent system for low-pressure die-cast process parameters optimization. *Int. J. Adv. Manuf. Technol.* 65, 517–524.
7. Zhao X, He Z, Zhang S, Liang D (2015) A sparse-representation-based robust inspection system for hidden defects classification in casting components. *Neurocomputing* 153: 1–10
8. Amit Sata, Ravi B (2017) Bayesian inference-based investment-casting defect analysis system for industrial application. *Int. J. Adv. Manuf. Technol.* 90:3301–3315
9. Galan U, Orta P, Kurfess T, Ahuett-Garza H (2018) Surface defect identification and measurement for metal castings by vision system. *Manuf. Lett.* 15: 5–8
10. Jeongsu Lee, Young Chul Lee, Jeong Tae Kim (2021) Migration from the traditional to the smart factory in the die-casting industry: Novel process data acquisition and fault detection based on artificial neural network. *Journal of Materials Processing Tech.* 290–116972
11. Niu B, Qu X, Guan X, Zhang F (2021) Rapid detection of highly reflective surface defects based on digital micromirror device. *Opt. Commun.* 501:127385
12. Yang Q, Jin J, Chang TS (2010) Automatic Feature Extraction and Classification of Surface Defects in Continuous Casting. *Transactions North American Manufacturing Research Institution of SME* 38: 563–70
13. Lin J, Yao Y, Ma L, Wang Y (2018) Detection of a casting defect tracked by deep convolution neural network. *Int. J. Adv. Manuf. Technol.* 97: 573–81
14. Ai Y, Xu K (2013) Surface Detection of Continuous Casting Slabs Based on Curvelet Transform and Kernel Locality Preserving Projections. *J. Iron Steel Res. Int.* 20 (5): 80–6
15. Yu Y, Luo X, Liu Q (2017) Model predictive control of a dynamic nonlinear PDE system with application to continuous casting. *J. Process Control* 65: 41–55
16. Bouhouche S, Yahi M, Bast J (2011) Combined use of principal component analysis and self-organization map for condition monitoring in pickling process. *Appl. Soft Comput.* 11: 3075–82
17. Bouhouche S., Laïb DLY, Bast J (2014) Evaluation-based closed-loop errors using principal component analysis and self-organisation map with an application to a pickling process. *Int. J. Adv. Manuf. Technol.* 70: 1033–41
18. Bendjama H, Mahdi D (2016) Computer Code for Materials Diagnosis Using Monte Carlo Method and Neural Networks. *J. Fail. Anal. Prev.* 16 (4): 931–37
19. Anderson TW (2003) An introduction to multivariate statistical analysis. California: John Wiley & Sons
20. Tamura M, Tsujita S (2007) A study on the number of principal components and sensitivity of fault detection using PCA. *Comput. Chem. Eng.* 31: 1035–46
21. Jackson JE (2003) A User's Guide to Principal Components. New York: John Wiley & Sons
22. Kohonen T (2001) Self-organizing Maps. Berlin: Springer-Verlag
23. Lang L (1999) Development and testing of a low pressure lost foam casting machine and an examination of the process. PhD thesis, Freiberg University, Germany
24. Bast J, Aitsuradse M, Hahn T (2004) Advantages of the Low Pressure Lost Foam Casting Process. *AFS Transactions* 112: 1131–1144