www.czasopisma.pan.pl PAN www.journals.pan.pl

ARCHIVES of FOUNDRY ENGINEERING

DOI: 10.2478/v10266-012-0127-5

Published quarterly as the organ of the Foundry Commission of the Polish Academy of Sciences

VERSITA

ISSN (2299-2944)
Volume 12
Issue 4/2012

171 – 175

Application of Time-series Analysis in Control of Chemical Composition of Grey Cast Iron

M. Perzyk*, A. Rodziewicz
Institute of Manufacturing Technologies, Warsaw University of Technology,
ul. Narbutta 85, 02-524 Warszawa, Poland
* Corresponding author. E-mail address: M.Perzyk@wip.pw.edu.pl

Received 06.06.2012; accepted in revised form 03.09.2012

Abstract

The aim of the paper was an attempt at applying the time-series analysis to the control of the melting process of grey cast iron in production conditions. The production data were collected in one of Polish foundries in the form of spectrometer printouts. The quality of the alloy was controlled by its chemical composition in about 0.5 hour time intervals. The procedure of preparation of the industrial data is presented, including OCR-based method of transformation to the electronic numerical format as well as generation of records related to particular weekdays. The computations for time-series analysis were made using the author's own software having a wide range of capabilities, including detection of important periodicity in data as well as regression modeling of the residual data, i.e. the values obtained after subtraction of general trend, trend of variability amplitude and the periodical component. The most interesting results of the analysis include: significant 2-measurements periodicity of percentages of all components, significance 7-day periodicity of silicon content measured at the end of a day and the relatively good prediction accuracy obtained without modeling of residual data for various types of expected values. Some practical conclusions have been formulated, related to possible improvements in the melting process control procedures as well as more general tips concerning applications of time-series analysis in foundry production.

Keywords: Application of Information Technology to the Foundry Industry, Quality Management, Time-series Analysis, Melting Process, Grey Cast Iron

1. Introduction

Time-series analysis is one of the data mining methods, which deals with series of data recorded in a chronological order, usually in regular time intervals or in another sequences. There are two aims of that kind of analysis: the discovery of the nature of the given process or phenomenon and prediction of future values. Time-series prediction can be considered as a particular case of a regression task, where the input and output variables are the same quantity but measured at different time moments.

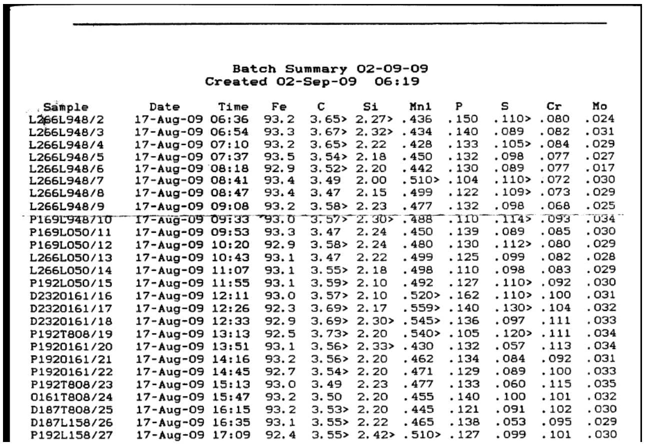
The time-series analysis has been widely applied in business problems, however, recently multiple examples of their utilization in technology, including manufacturing industry, can be found in literature (e.g. [1-5]). As indicated in [6], the application of these methods in foundry industry can bring essential benefits. Detection of periodicity in products or materials properties manufactured in a foundry can facilitate identification of the irregularities in the manufacturing processes. Prediction of the process parameters or the product properties can help to prevent unwanted tendencies or suggest required changes in the process control.

2. Methodology of the research

2.1. Industrial data preparation

In the present work the time-series analysis was applied to the melting process of grey cast iron, run in one of the Polish foundries. The melting process was controlled by the chemical composition of the iron, defined by 8 elements (Fe, C, Si, Mn, P, S, Cr, Mo). The percentages of the elements were measured by a spectrometer and available in a form of printouts. These values were recorded in irregular time intervals, ranging from 10 to 55 minutes, with average about 25 minutes. All the data were collected during 3 months (July, August and September 2009) and contained about 1700 records.

The original printed data were transformed to the electronic format using an OCR (Optical Character Recognition) software ABBYY FineReader 9.0, provided with an Epson scanner. A sample of the original printout is shown in Fig. 1. The conversion of one A4 page of the printout takes about 1 minute. The percent of the correctly recognized characters exceeds 99,9%. After the conversion to the text type computer file, the data was imported to Excel and the faulty characters were found and 'manually" corrected.



Batch Summary 02-09-09
Created 02-Sep-09 06:19

Sample	Date	Time	Fe	C	Si	Mn1	P	S	Cr	Mo
L266L948/2	17-Aug-09	06:36	93.2	3.65>	2.27>	.436	.150	.110>	.080	.024
L266L948/3	17-Aug-09	06:54	93.3	3.67>	2.32>	.434	.140	.089	.082	.031
L266L948/4	17-Aug-09	07:10	93.2	3.65>	2.22	.428	.133	.105>	.084	.029
L266L948/5	17-Aug-09	07:37	93.5	3.54>	2.18	.450	.132	.098	.077	.027
L266L948/6	17-Aug-09	08:18	92.9	3.52>	2.20	.442	.130	.089	.077	.017
L266L948/7	17-Aug-09	08:41	93.4	3.49	2.00	.510>	.104	.110>	.072	.030
L266L948/8	17-Aug-09	08:47	93.4	3.47	2.15	.499	.122	.109>	.073	.029
L266L948/9	17-Aug-09	09:08	93.2	3.58>	2.23	.477	.132	.098	.068	.025
P169L948/10	17-Aug-09	09:33	93.0	3.57>	2.30>	.488	.110	.114>	.093	.034
P169L050/11	17-Aug-09	09:53	93.3	3.47	2.24	.450	.139	.089	.085	.030
P169L050/12	17-Aug-09	10:20	92.9	3.58>	2.24	.480	.130	.112>	.080	.029
L266L050/13	17-Aug-09	10:43	93.1	3.47	2.22	.499	.125	.099	.082	.028
L266L050/14	17-Aug-09	11:07	93.1	3.55>	2.18	.498	.110	.098	.083	.029
P192L050/15	17-Aug-09	11:55	93.1	3.59>	2.10	.492	.127	.110>	.092	.030
D2320161/16	17-Aug-09	12:11	93.0	3.57>	2.10	.520>	.162	.110>	.100	.031
D2320161/17	17-Aug-09	12:26	92.3	3.69>	2.17	.559>	.140	.130>	.104	.032
D2320161/18	17-Aug-09	12:33	92.9	3.69>	2.30>	.545>	.136	.097	.111	.033
P192T808/19	17-Aug-09	13:13	92.5	3.73>	2.20	.540>	.105	.120>	.111	.034
P1920161/20	17-Aug-09	13:51	93.1	3.56>	2.33>	.430	.132	.057	.113	.034
P1920161/21	17-Aug-09	14:16	93.2	3.56>	2.20	.462	.134	.084	.092	.031
P1920161/22	17-Aug-09	14:45	92.7	3.54>	2.20	.471	.129	.089	.100	.033
P192T808/23	17-Aug-09	15:13	93.0	3.49	2.23	.477	.133	.060	.115	.035
0161T808/24	17-Aug-09	15:47	93.2	3.50	2.20	.455	.140	.100	.101	.032
D187T808/25	17-Aug-09	16:15	93.2	3.53>	2.20	.445	.121	.091	.102	.030
D187L158/26	17-Aug-09	16:35	93.1	3.55>	2.22	.465	.138	.053	.095	.029
P192L158/27	17-Aug-09	17:09	92.4	3.55>	2.42>	.510>	.127	.099	.101	.030

Fig. 1. A sample of the original printout of the chemical composition testing obtained from the foundry spectrometer

Two types of data records were prepared. The first one included all the records available and was aimed at analysis of short-term variability in the data, including detection of possible periodicity and evaluation of the prediction capability of the time-series analysis.

The other type of records was oriented to detection of the possibly variability and periodicity related do the day of week. Because not all the working week-days were equally represented in the data (due to holidays and other breaks in the production), it was possible to find only 16 days representing four repeatable sequences of four week days, from Monday to Thursday. The following types of the daily based values were taken as the 16-records data sets:

- first measurement of the day,
- average of the first 3 measurements of the day,
- average of the last 3 measurements of the day,
- average of all measurements of the day.

All the data were prepared separately for all eight chemical elements measured by the spectrometer, resulting in total number 40 data sets.

2.2. Methodology of time-series analysis

The analysis and prediction of time –series can be done by many different methods. Time-series models have three classical types: Auto Regressive (AR), Integrated) (I) and with Moving Average (MA). The compositions of those three classes make the popular autoregressive with moving average models (ARMA) as well the autoregressive integrated with moving average (ARIMA). An alternative is application for a time-series a generalized regression model.

In the present work the latter approach is applied, described in detail in [6]. The idea is to utilize a multivariate regression model (in the present work it was simple linear regression) in which the input variables are values of the given quantity recorded in consecutive moments, and the output variable is its value shifted by one or several measurements from the last of the input points. On the basis of the known values in the time-series data, the data records for regression modeling are created, by taking the above mentioned input – output sets, shifted by one point in the series. The composition of the records for regression model is preceded by subtraction of the general trend, the variability amplitude trend as well as the periodical component. The idea of this methodology is to use a regression model for modeling finer changes than those which can be easily described by trends and periodicity.

The above procedure was implemented in a software created in the Institute of Manufacturing Technologies of Warsaw University of Technology. The program uses MS Excel platform and the VBA as programming language. The most significant periodical component is always subtracted, irrespective its statistical significance.

The other details concerning the theoretical basis of the procedures included in the software can be found in [6] and the works cited there.

3. Results

3.1. Periodicity in the data

For all seven data sets containing all recorded values in chronological order the periodicity equal 2 appeared to be significant. This implies that the most significant type of the variability was "up and down". This result does not evident from the observations of the data presented in the graphical form (see a typical graph in Fig. 2). That kind of behavior, combined with a relatively large amplitude of the variability, may indicate that some "overbalances" are present in the process control.

An additional analysis was made in which the growths and drops of all possible pairs of the elements were compared and the cases in which both changes go in the same directions were counted. The result was that in the 65% of the cases the changes of chromium and molybdenum contents were consistent. These elements often occur together, e.g. in alloyed steel scrap, and the observed periodicity may be a result of an imprecise proportioning of that kind of additives.

For the data sets referring to days of week, only in one case the periodicity appeared to be significant: the average Si content in the last 3 measurements of the day. Its value was equal 7. The interpretation of this result is not obvious or simple. For example, it could indicate that the stock of the additives containing silicon is suitable just for a period of about 7 working days and the practice of the operators is to spend it in full in that period (increasing or decreasing the amounts added at the end of the period).

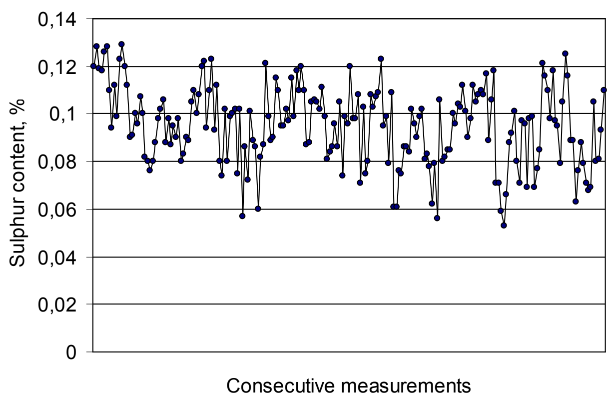


Fig. 2. Exemplary fragment of original data

3.2. Prediction accuracy

Essentially, the predictions of future values based on previous time-series data can be made by regression modeling of the residual values, as described in the previous section. For comparison, all the computations were also made without modeling the residual values, i.e. based on the both types of trends and periodicity. It is important that the most significant periodical component was calculated and taken into account in the computation even if it was not statistically significant (see Section 3.1).

In principle, the results of modeling of the residual data should be valuable if the information content in that data is significant. The software used in the computations provides that kind of information in a verbal form, based on results of two statistical tests: the runs test (also called Wald-Wolfowitz test) and the Durbin-Watson test, both described in [6] and the literature cited there. For the purposes of present work, these messages were converted to a numerical score scale, as shown in Table 1.

Table 1.
Possible messages concerning the residual data information content in time-series analysis, appearing in the applied software

Information displayed by the software used for time-series analysis	Conventional score values for information content in residual data
Residual data are only a noise and do not contain any significant information	1
Residual data are rather a noise and do not contain any significant information	2
Residual data may be not only a noise and may contain significant information	3
Residual data are rather not only a noise and contain significant information	4
Residual data are not only a noise and contain significant information	5

All the prediction errors were calculated as relative ones, i.e. the absolute values of differences between the predictions and the real values divided by the range of variability of a given variable (element content) in the time-series original data, ignoring the outliers.

In the present work the predictions for both types of time-series records described in Section 2.1 were made. For the daily based data sets, each comprising 16 records, the computations were made for all the chemical elements contents. Because of the small quantity of the available training records, the reliable regression modeling was not possible and all the predictions were made based on the both types of trends and periodicity only. In Fig. 3 exemplary daily-based data together with the corresponding residual values are shown.

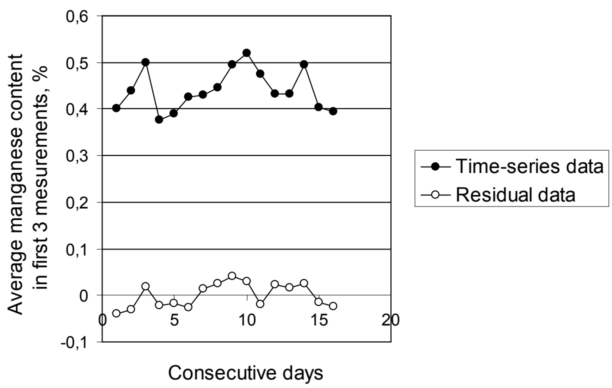


Fig. 3. Manganese percentage data for the average of the first 3 measurements of the day, appearing in the four sequences of 4 working days (Monday to Thursday) weeks

In Fig. 4 the prediction errors for the daily-based data, each averaged over all the 7 testing points, are shown. The lowest error values are for the whole day averages and are between 6 and 15%.

For the data including consecutive measurements (analysis of short-term variability in the data) the prediction accuracy of manganese and sulphur contents was evaluated for the points selected in the following way. Wednesdays were chosen as the working days with the largest numbers of measurements. The values recorded before 12 am were used as training data, in which the 5 consecutive measurements were inputs and the 6th one was an output in the linear regression models implemented in the software. The predictions for Mn and S contents were made, for the first measurements after 12 am in the following days: 1, 8, 15, 22 and 29 July, 19 and 26 August as well as 2, 9, 16, 23 and 30 September, 2009.

In Fig. 5 the relative prediction errors for the individual measurements described above are shown. The general level of accuracy of the predictions made without modeling residual data can be evaluated as satisfactory.

For the daily-based data (Fig. 4) the relative errors do not exceed 25% and their average value is less then 15%. The lowest errors, between 6 and 15%, were obtained for the averaged measurements over a day. It means that the planners would be able to instruct the operators to make some corrections of the typical amounts of additives for the next day.

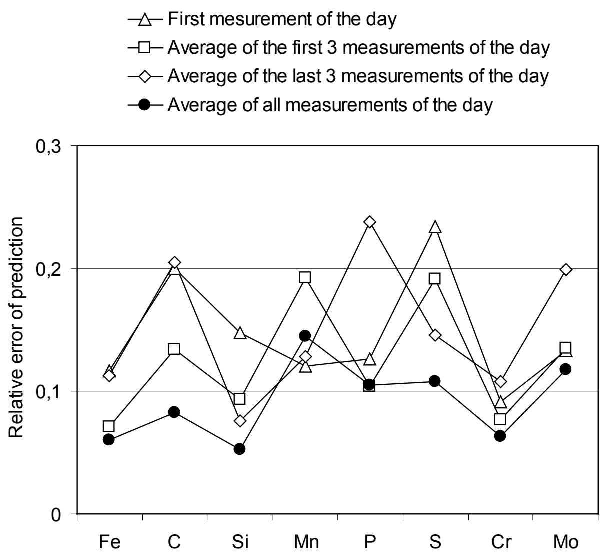


Fig. 4. Mean prediction errors for all the daily based time-series data sets (all types of day measurements and all chemical elements)

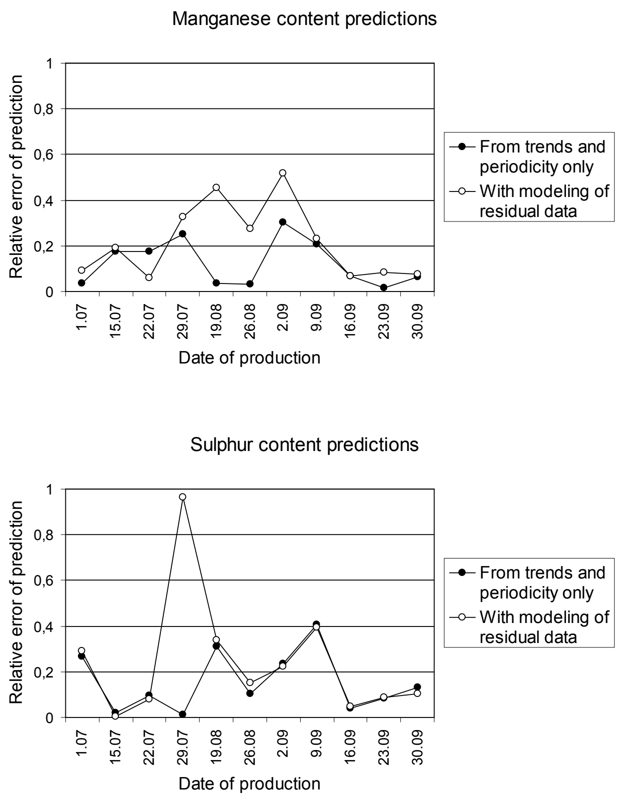
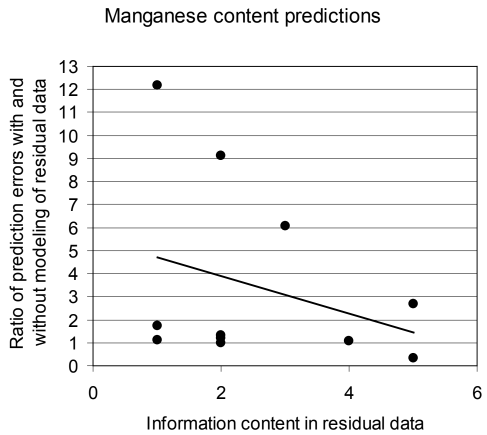


Fig. 5. Comparison of prediction errors for Mn and S contents in cast iron, for the first measurements recorded after 12 am, on Wednesdays, during three months

For the current content of Mn and S (Fig. 5) the relative errors do not exceed 40% and their average value is less then 20%. For the industrial melting process it means that, with a great confidence, the operators would be able to predict weather the next value of a given chemical element content will be high, medium or low. This would enable them to adjust the current amounts of the additives with a greater accuracy than if they rely only on the current results of chemical analysis.

In Fig. 6 the ratios of the prediction errors shown in Fig. 5 with modeling of residual data to those obtained from the both types of trends and periodicity, are presented as a function of the information contents in the residual data, according Table 1.



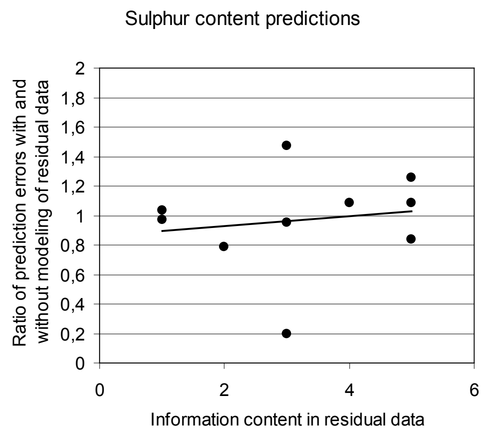


Fig. 6. Influence of the information contents in residual data on the effectiveness of regression modelling, data as in Fig 5

The results plotted in Fig. 6 indicate that the linear regression modeling of the residual data may not necessary improve the prediction accuracy. For the Mn content predictions, the residual data modeling increased the prediction errors in most of the cases, compared to those based only on the both types of trends and periodicity. This can be possibly attributed to the low information contents in the residual data; the increase of the information content apparently reduces this negative effect. For the S content predictions the effect of the residual modeling on the prediction accuracy is not so explicit.

4. Conclusions

The results obtained in the present work indicate that the time-series analysis can be a valuable data mining tool for detection important characteristics in production data. The periodicity values can point at some imperfections or abnormalities appearing in the production control, including those resulting from bad habits or organisational issues. However, the particular conclusions can be formulated by the engineering or operational staff.

The main purpose of the predictions of the future values in a production process is to enable the staff to take appropriate actions anticipating the expected changes in the process. The prediction accuracy for both current and daily-based averaged values of the chemical component of grey cast iron appeared to be satisfactory for that purpose.

It is recommended to use the residual data modelling with caution, probably only in the cases when the information content in that data is definitely significant.

Future research would be desirable in order to analyse the behaviour of the regression models for residual data, including more advanced, non linear ones, e.g. artificial neural networks or regression trees. Also testing the methodology of time-series analysis used in the present work using a large variety of data as well the comparison with ARIMA type analysis would be advisable.

Acknowledgment

This work was partly supported by project No. N R07 0015 04 granted by Ministry of Science and Higher Education, Poland.

References

[1] Bisgaard, S. & Kulahci, M. (2008). Quality quandaries: Using a time series model for process adjustment and control. *Quality Engineering*, 20, 134–141. DOI: 10.1080/08982110701771042.

[2] Lin, Z.C. & Chang, D.Y. (2009). A band-type network model for the time-series problem used for IC leadframe dam-bar shearing process. *International Journal of Advanced Manufacturing Technology*. 40 (11-12), 1252-1266. DOI: 10.1007/s00170-008-1432-6.

[3] Kapoor, S.G. & Terry, R.W. (1983). Time series analysis of tensile strength in a die casting process, Time Series Analysis: Theory and Practice 4, Proc. 8th International Time Series Meeting and 3rd American Conference, 1983, 221-228.

[4] Camisani-Calzolari, F.R., Craig, I.K. & Pistoriust, P.C. (2003). Quality prediction in continuous casting of stainless steel slabs, *Journal of The South African Institute of Mining and Metallurgy*, 103 (10), 651-665.

[5] Perzyk, M., Kochański, A., Biernacki, R., Kozłowski, J. & Soroczyński, A. (2011). Applications of data mining to diagnosis and control of manufacturing processes, in "Knowledge – Oriented Applications in Data Mining", eds. Kimito Funatsu and Kiyoshi Hasegawa, Publ. InTech, 2011, 147-166. DOI: 10.5772/13282.

[6] Perzyk, M., Krawiec, K. & Kozłowski, J. (2009). Application of time-series analysis in foundry production, *Archives of Foundry Engineering*. 9 (3), 109-114.