ARCHIVES
of
FOUNDRY ENGINEERING

ISSN (1897-3310)
Volume 9
Issue 1/2009
9-12

2/1

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

Simulation by the method of inverse cumulative distribution function applied in optimising of foundry plant production

J. Szymszal [a] *, J. Piątkowski [a], P. Podolski [a]., J. Kliś [b]
[a] Chair of Metal Alloys and Composites Engineering, Silesian University of Technology,
ul. Krasińskiego 8, 40-019 Katowice, Poland
[b] Centrostal Górnośląski sp. z o.o., ul. Stalowa 1, 40-610 Katowice, Poland
*Contact for correspondence: e-mail: jan.szymszal@polsl.pl

Received 26.02.2009; accepted in revised form: 30.03.2009

Abstract
The study discusses application of computer simulation based on the method of inverse cumulative distribution function. The simulation refers to an elementary static case, which can also be solved by physical experiment, consisting mainly in observations of foundry production in a selected foundry plant. For the simulation and forecasting of foundry production quality in selected cast iron grade, a random number generator of Excel calculation sheet was chosen. Very wide potentials of this type of simulation when applied to the evaluation of foundry production quality were demonstrated, using a number generator of even distribution for generation of a variable of an arbitrary distribution, especially of a preset empirical distribution, without any need of adjusting to this variable the smooth theoretical distributions.

Keywords: computer-aided foundry production, simulation, quality, method of inverse cumulative distribution function

1. Introduction

Simulation means using a set of different research techniques, based on activation of a model imitating the real behaviour of the examined system. Owing to the simulation we are equipped with a tool that enables us to observe various phenomena and examine them in a way that until now has been possible only in true experiment. In this study, an elementary static case has been presented, which can also be solved by physical experiment. Because of large differentiation in models of the discrete events, their theoretical representation requires high level of abstraction, and for this reason, to demonstrate operation of such models, a specific example based on simulation using the method of inverse cumulative distribution function has been chosen. A solution of this type is expected to help in determination of the profitability of foundry production.

2. Problem description

A foundry plant makes cast iron, which serves as a starting material for the manufacture of more complex grades. The cast iron satisfies the technical conditions of acceptance, if (among others) it has the required chemical composition. To make the example simpler, only the three main alloying elements were included in the analysis, i.e. carbon, silicon and manganese. Their content should amount to: C% from 2,9 to 3,3%, Si% from 1,2 to 1,4%, Mn% from 0,30 to 0,60%. For this foundry, the production will be profitable, if at least one half of the melted iron satisfies

the quality imposed by the standard, i.e. if the required chemical composition is simultaneously obtained in respect of the C%, Si% and Mn% content.

3. Physical experiment – production

Since some time, the foundry plant has been manufacturing the examined cast iron, and as a result of physical experiment, the following real probability distributions of the content of the alloying elements with breakdown into classes (given in Table 1) were recorded.

Table 1.
The distributions and cumulative distribution functions of the content of C%, Si% and Mn% with breakdown into classes

Carbon (C%)			Silicon (Si%)			Manganes (Mn%)		
C%	Prob.	Distrib.	Si%	Prob.	Distrib.	Mn%	Prob.	Distrib.
2,8 - 2,9	0,06	0,06	1,0 - 1,2	0,09	0,09	0,20 - 0,30	0,08	0,08
2,9 - 3,0	0,20	0,26	**1,2 - 1,4**	0,85	0,94	**0,30 - 0,40**	0,22	0,30
3,0 - 3,1	0,33	0,59	1,4 - 1,6	0,06	1,00	**0,40 - 0,50**	0,39	0,69
3,1 - 3,2	0,26	0,85				**0,50 - 0,60**	0,24	0,93
3,2 - 3,3	0,10	0,95				0,60 - 0,70	0,07	1,00
3,3 - 3,4	0,05	1,00						

Since a large batch of the melts has been examined, it can be assumed that the results describe with sufficient accuracy the chemical composition of the cast iron in general population, which means that the empirical frequencies of occurrence of the individual classes of the content of alloying elements can be regarded as a probability of their occurrence. Additionally, in Table 1, the cumulative values of the real probability (the cumulative distribution function) were computed.

4. Sampling by the method of inverse cumulative distribution function

The main inconvenience in practical solution of the problem lies in the fact that it requires physical production of the examined cast iron, and if the chemical composition demanded by respective standards is not complied with, this solution may prove both costly and invalid. From the assumptions adopted in physical experiment (related e.g. with the central limit theorem) it follows, however, that there is no need to continue the physical experiment to know if the production of cast iron can be profitable under certain conditions. It is enough to sample at random the variable values of C%, Si% and Mn%. It is also important to ensure that the probability distributions of the sampled variables are consistent with the corresponding distributions in general population. If so, then it can be assumed that the C%, Si% and Mn% variables are sampled from these distributions. Every next repetition of the sampling operation (also known as replication) provides the next observation on the variables of C%, Si% and Mn%, and consequently on the cumulative cast iron composition. And so, e.g., the cast iron containing carbon (C%) in a range of 2,9 – 3,1% should appear in the sample 20 times out of hundred, while the cast iron alloyed with manganese (Mn%) comprised in a range of 0,50 – 0,60% should appear 24 times out of hundred.
For random sampling of the C%, Si% and Mn% variables from the respective distributions, a variable taking with equal chances the values from an interval (0,1), i.e. the variable of a uniform (even) distribution along the line segment (0,1), should be applied. Let a variable of this type, called random variable, be denoted by the symbol U. In an Excel calculation sheet it can be generated using a non-argument function =LOS(). Figure 1 shows possibilities of using this function in generation of numbers from the selected interval and distribution.

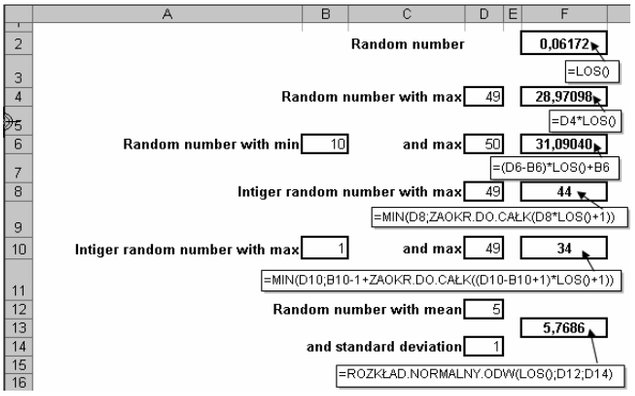

Fig. 1. Application of function =LOS() for generation of numbers from the selected interval and distribution

The following rules have been adopted:

- If the value of the random variable U1 is not larger than 0,06, the content of C% is comprised in the class of 2,8 – 2,9%, otherwise
- If the value of the random variable U1 is not larger than 0,26 the content of C% is comprised in the class of 2,9 – 3,%, otherwise
- If the value of the random variable U1 is not larger than 0,59 the content of C% is comprised in the class of 3,0 – 3,1%, otherwise
- If the value of the random variable U1 is not larger than 0,85 the content of C% is comprised in the class of 3,1 – 3,2%, otherwise
- If the value of the random variable U1 is not larger than 0,95 the content of C% is comprised in the class of 3,2 – 3,3%, otherwise
- the content of C% is comprised in the class of 3,3 – 3,4%.

So, it follows that the sampled variable C% has the carbon content comprised in a class corresponding to the line where the cumulative distribution function for the first time reaches or exceeds the value of the random variable. A similar principle has been adopted for the remaining two random variables, i.e. Si% and Mn%, referring them to proper distributions in general population. Figure 1. shows, in an Excel sheet, 1000 replicates using the random variables U1 (C%), U2 (Si%) and U3 (Mn%).

5. The statistical analysis

The test shown in Figure 1 is subject to statistical processing. In this case, as appropriate, for all the three variables, i.e. C%, Si% and Mn%, the main characteristics of descriptive statistics, i.e. the mean, standard deviation, median, 25% and 75% quartiles, and modal value, were computed.

For each analysed variable, using table function

=(FREQUENCY), the frequency of occurrence of a given class of the alloying element content (column: simulated frequency (ni)) was determined, and basing on these data, the cumulative frequency and the simulated probability were computed (Fig. 2).

The correctness of approach used in this method can be verified further by evaluation of consistency between the empirical distribution and simulated distribution, using for this purpose a chi-square test, performed with function =TEST.CHI(). In each case (C%, Si%, Mn%), a very good consistency between the two distributions was obtained.

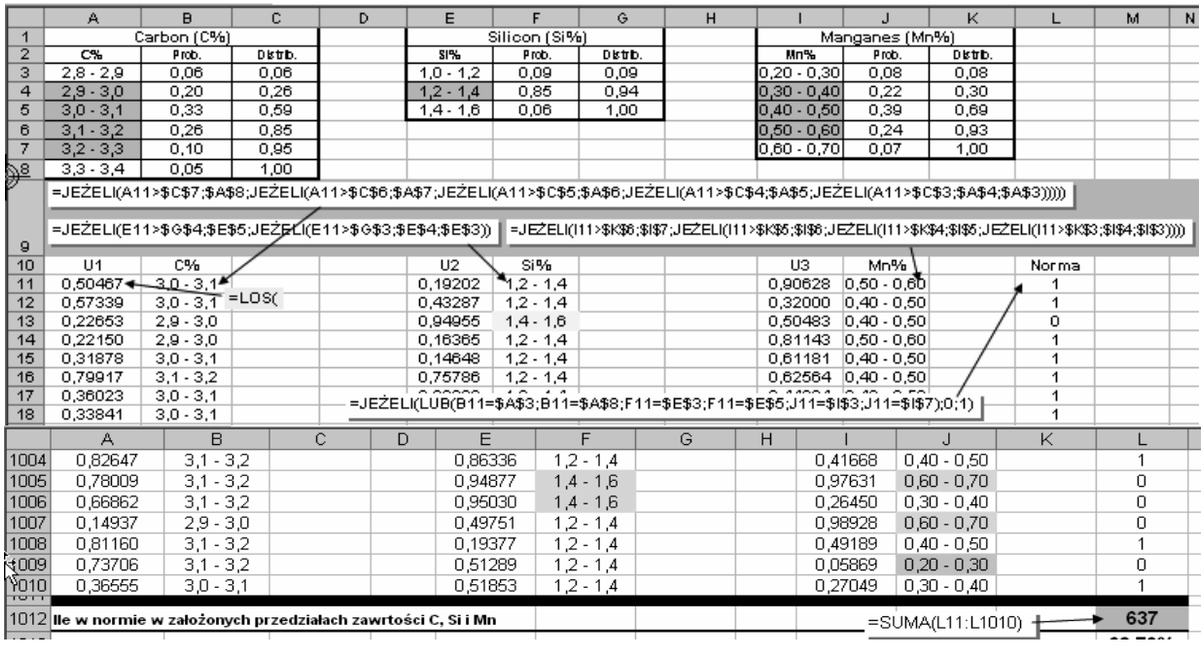

Fig.1. Sheet with the results of simulation done by the method of inverse cumulative distribution for 1000 replications

a) b) c)

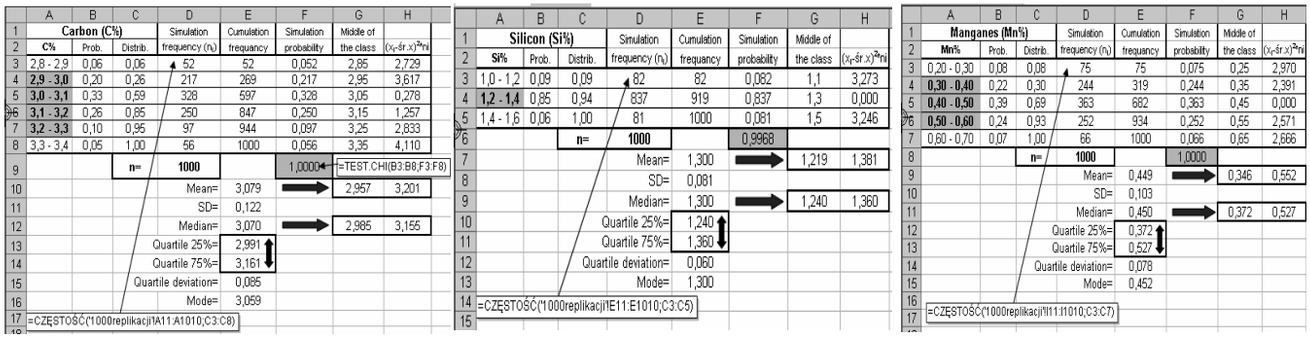

Fig. 2. Estimated fundamental characteristics of the descriptive statistics for a) carbon; b) silicon; c) manganese

As a first step, for each of the examined variables (C%, Si%, Mn%), presented in the form of a stemplot, the arithmetic mean was estimated according to the following formula:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} \dot{x}_i \cdot n_i \qquad (1)$$

where: $\dot{x}_i$ – the middle of a class interval

k – the number of classes

n –the total size

Having computed the difference between the middle of class intervals and the arithmetic mean, and raising it to the second power, the value of the standard deviation s was computed according to the following formula:

$$s = \sqrt{\frac{1}{n}\sum_{i=1}^{k}\left(\dot{x}_i - \bar{x}\right)^2 \cdot n_i} \qquad (2)$$

The values of the mean and of a standard deviation *s* obtained for the individual variables enabled the determination of a typical variability interval, which should comprise 68% of all melts.
As a next step, the median, i.e. the middle value, was estimated using the following formula:

$$Me = x_{0Me} + \frac{N_{Me} - \sum_{i=1}^{Me-1} n_i}{n_{Me}} \cdot h_{Me} \tag{3}$$

where:

$\sum_{i=1}^{Me-1} n_i$ - the cumulative size of an interval preceding the median,

$n_{Me} = n/2$ - the position of the median,

x_{0Me} - the lower limit of an interval where the modal value,

n_{Me} - the size of an interval where the median is present,

h_{Me} - the range of a class interval where the median is present.

The values of the quartiles: lower Q_1 (25%) and upper Q_3 (75%) were determined from the following formulae:

$$Q_1 = x_{0Q1} + \frac{N_{Q1} - \sum_{i=1}^{Q1-1} n_i}{n_{Q1}} \cdot h_{Q1} \tag{4}$$

$$Q_3 = x_{0Q3} + \frac{N_{Q3} - \sum_{i=1}^{Q3-1} n_i}{n_{Q3}} \cdot h_{Q3} \tag{5}$$

where:

$N_{Q1} = \frac{n}{4}$, whereas $N_{Q3} = \frac{3n}{4}$

$n_{Q1}; n_{Q3}$ - the size of an interval where the examined quartile,

$\sum_{i=1}^{Q1-1} n_i; \sum_{i=1}^{Q3-1} n_i$ - the cumulative size of an interval preceding the examined quartile,

$h_{Q1}; h_{Q3}$ - the range of an interval where the examined quartile.

As regards the examined variables, a half of the melts (50%) is comprised within the interquartile range. Additionally, the computed values of the quartiles enable us to compute the quartile deviation Q, first:

$$Q = \frac{Q_3 - Q_1}{2} \tag{6}$$

and the typical variable range, next:

$$Me - Q < x_{typ} < Me + Q \tag{7}$$

As a last step, the modal value *Mo* is computed:

$$Mo = x_{0m} + \frac{n_m - n_{m-1}}{(n_m - n_{m-1}) + (n_m - n_{m+1})} \cdot h_m \tag{8}$$

where:

x_{0m} - the lower limit of an interval where the modal value

n_m - the size of the modal interval,

$n_{m-1}; n_{m+1}$ - the size of classes preceding and following the modal interval,

h_m - the range of a class interval where the modal value.

6. Summing up and conclusions

Commonly, the cumulation is done with standard measures based on sample moments of the variables. In the case under discussion, the mean values (after 1000 replications) obtained for all the examined variables (C%, Si%, Mn%) fully corresponded to the preset values. With the distribution of these variables close to normal, one might expect that about 68% of all melts would be comprised in the intervals determined by a variability interval typical of the mean value and standard deviation. Yet, this is not a satisfactory information, because the set of events defining a correct melt is composed of the sets of events operating in an independent mode for each of the examined alloying elements.

More precise information on the profitability of production is provided by the measures of position, like the median and quartiles. A rough estimation of these characteristics has proved that all the examined variables were characterised by the values comprised in an interquartile range, which means that, for the examined variables (C%, Si%, Mn%), at least 50% of the melts should satisfy the preset conditions.

To adapt the technique used for processing of the simulation results to the key question posed by every foundry plant: “how much”, one might disregard the previously estimated parameters of descriptive statistics and reduce the analysis to the count of melts corresponding to standards adopted previously. An answer to the question (column L – Fig. 1) shows that as many as 63,7% of all melts satisfy these conditions. Even if these results are burdened with some sampling variability, an adequately large margin of safety is still left. Some attention deserves the fact that the above measures were determined basing on the same measure that was used in replications. Their usefulness to the foundry plant is determined by the question that the cast iron manufacturer should ask himself. The answer most useful is obtained when the procedure used for processing of the simulation results (counting of melts that raise the profitability of production) is well adapted to a problem the decision-maker is currently facing.

References

[1] Gajda J.: Forecasting and simulation vs economic decisions, Wydawnictwo C.H. Beck, Warszawa 2001 (in Polish).
[2] Stasiewicz S., Rusnak Z., Siedlecka U.: Statistics – elements of the theory and tasks, Ed. by Akademia Ekonomiczna in Wrocław, Wrocław 1999 (in Polish).
[3] Winston W.L., Albright S.Ch.: Practical Management Science, Duxbury, Belmont 1997.
[4] Statistical data of selected Iron Foundry (in Polish).