Center for Quality and Productivity Improvement
UNIVERSITY OF WISCONSIN
610 Walnut Street
Madison, Wisconsin 53705

(608) 263–2520
(608) 263–1425 FAX

Report No. 50

An Application of Box–Jenkins Methodology to the Control of Gluten Addition in a Flour Mill

T. Fearn and P. I. Maris

August 1990

C Q P I
UNIVERSITY OF WISCONSIN

Report No. 50

Report Series
in Quality and Productivity

An Application of Box–Jenkins Methodology to the Control of Gluten Addition in a Flour Mill

T. Fearn
University College
London, UK

P. I. Maris
Flour Milling and Baking
Research Association
Chorleywood, UK

ABSTRACT

The approach of Box and Jenkins was used to design a control algorithm for a feedback loop controlling the addition of dried gluten to breadmaking flour in a flour mill. The variations to be controlled were modelled by an IMA(0,1,1) process and the system dynamics identified as a simple delay. The resulting optimal control strategy was implemented and worked well.

KEYWORDS: *ARIMA, Box–Jenkins, flour milling, gluten, near-infrared, optimal control, stochastic control, time series*

An Application of Box–Jenkins Methodology to the Control of Gluten Addition in a Flour Mill

T. Fearn and P. I. Maris

The approach of Box and Jenkins was used to design a control algorithm for a feedback loop controlling the addition of dried gluten to breadmaking flour in a flour mill. The variations to be controlled were modelled by an IMA(0,1,1) process and the system dynamics identified as a simple delay. The resulting optimal control strategy was implemented and worked well.

1. INTRODUCTION

Box and Jenkins (1976) present a unified approach to forecasting and control problems based on the modelling of time series by autoregressive–integrated–moving–average (ARIMA) processes. There are many published examples of this approach in the areas of modelling and forecasting, but there seem to be fewer in control, in the statistical literature at least.

This paper describes the successful use of the Box–Jenkins methodology to design a control algorithm for a feedback loop controlling the addition of dried gluten to breadmaking flour in a flour mill.

2. CONTROL OF GLUTEN ADDITION

2.1 GLUTEN ADDITION

Flour milled from home-grown or other European Community (EC) wheat rarely contains enough protein to make high volume bread of the sort popular in the UK. One solution commonly adopted is for the miller to add extra protein in the form of dried gluten powder. This is made by washing most of the starch out of wheat flour and usually contains around 75% natural wheat protein. A typical case might involve the addition of 2% dried gluten by weight to a flour of original protein content 10%, thus raising the protein to 11.3%.

The dried gluten is usually added at the end of the milling process by a feeder that can be set to add, for example, 120 kg/hr to a flour stream of 6000 kg/hr. Depending on the grade of flour being produced there will be a target protein content for the flour. It is important that the protein content does not fall below this target, for then the flour will not perform to specification. Gluten is expensive so it is also important not to add more than is necessary. Add to this the fact that the protein content of the flour stream varies with time and the need for a control system is evident.

2.2 AN AUTOMATIC CONTROL SYSTEM

Figure 1 shows the automatic control system for gluten addition developed by the Flour Milling and Baking Research Association (FMBRA), as installed

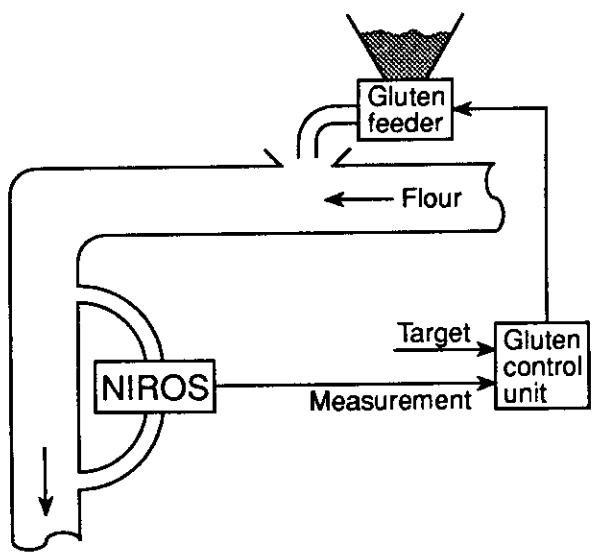

Figure 1. *Schematic representation of the automatic gluten control system.*

in a commercial flour mill. The key element in the system is the box labelled NIROS. This is an automatic sampling device attached to a near infrared spectrophotometer. It samples flour from the main production stream once per minute, measures the protein content, and sends the result to a remote box of electronics, the gluten control unit. Here the measurement is compared with the target and an adjustment made to the set point of the gluten feeder with the aim of keeping the flour protein content on target.

The rule used is to change the gluten addition rate after each new measurement by an amount calculated to correct for one quarter of the discrepancy between measurement and target. The deviation of this rule using the approach of Box and Jenkins is described in Section 3.

Feedback control was chosen in preference to feedforward control because a feedback system is intrinsically safer. Measuring after the addition point provides a record of the protein contents actually achieved and makes it possible to detect any malfunctions. A feedback system is also more robust. For example a 10% calibration error in the gluten feeder would result in a slight loss of efficiency rather than a systematic error in protein levels. This point is returned to in Section 5.

3. DERIVATION OF CONTROL ALGORITHM

Box and Jenkins derive optimal control schemes by combining a stochastic model for the variations to be controlled and a dynamic model for the system response to control actions. In this case we need to know how protein levels vary with time and what happens when the setting on the gluten feeder is changed.

3.1 MODELLING THE VARIATION IN PROTEIN CONTENT

Twelve series of protein measurements were analyzed. All of the series were collected by the NIROS instrument at a sampling rate of one per minute; they varied in length from 107 minutes to 766 minutes. Series 1–5 were available at the beginning of the work, having been collected in another mill at an earlier stage of the NIROS project. Series 6–12 were collected in the mill where the control system was installed, after the installation of NIROS but before the completion of the control system. All the series represent the "uncontrolled" variation in flour protein content. Gluten was added to some of the flours, but at a constant level. Series 2, the longest, is shown in Figure 2.

For technical reasons the protein measurements were only recorded to one decimal place, although it is obvious from the plot that two would have been more appropriate. The precise effect of this on the analysis of the series is not clear; it seems unlikely however that the conclusions would have been qualitatively different had more precise data been available.

To identify a model the sample autocorrelation and partial autocorrelation functions for the original series and their first and second differences were

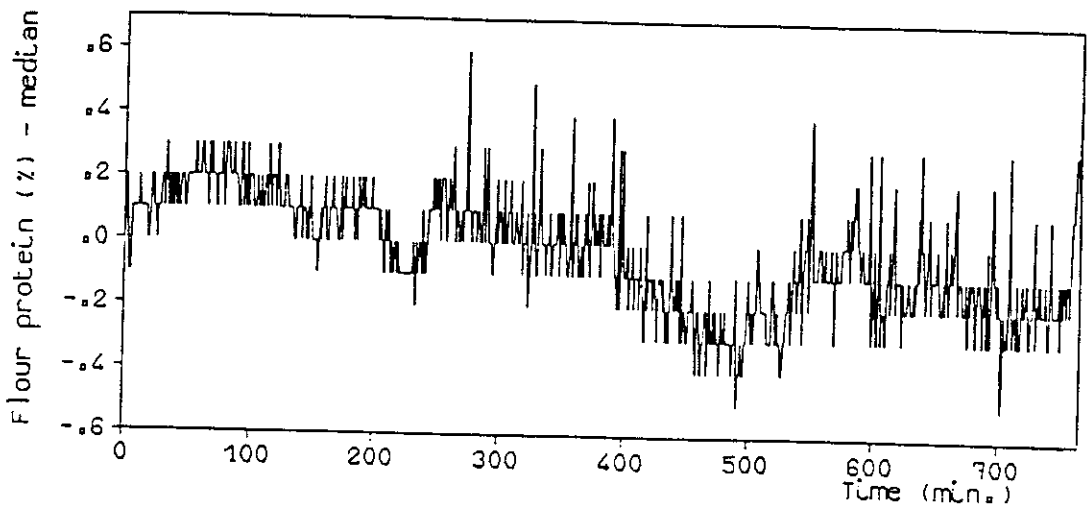

Figure 2. *Series of 766 measurements of flour protein made by the NIROS instrument at a rate of one per minute (Series 2). The median protein content has been subtracted from all measurements.*

examined (Figure 3). In each case the model clearly indicated was IMA(0,1,1), a first order moving average process for the first differences of the protein measurements, i.e.

$$z_t - z_{t-1} = a_t - \theta a_{t-1}$$

where z_t is the series of protein measurements, a_t is a series of independent N(0, σ_a^2) disturbances, and θ is the moving average parameter. Estimating θ and σ_a for the twelve series gave the results in Table 1.

Table 1
Lengths and standard deviation of 12 series of protein measurements and results of fitting the IMA(0, 1, 1) *model*

			Estimates of			
series	*length*	*standard deviation*	θ	SE(θ)	σ_a	Q^*
1	290	.13	.752	.039	.11	22.0
2	766	.17	.810	.023	.11	25.4
3	346	.14	.908	.023	.12	15.4
4	280	.09	.686	.046	.07	24.8
5	563	.10	.898	.022	.10	31.5
6	170	.15	.906	.035	.14	29.0
7	314	.22	.800	.034	.15	17.1
8	107	.30	.790	.061	.19	17.2
9	280	.14	.648	.046	.14	13.5
10	160	.16	.860	.042	.14	23.8
11	180	.18	.766	.049	.14	18.5
12	209	.19	.890	.032	.15	33.8

*Q is the modified Box–Pierce statistic calculated from the first 20 autocorrelations of the series of residuals.

To investigate the fit of the model, the sample autocorrelation function of the residuals for each series was inspected and the modified Box–Pierce statistic (Ljung and Box, 1978) was calculated from the first twenty autocorrelations. The values of this test statistic, whose distribution is approximately chi-squared on nineteen degrees of freedom if all the true autocorrelations are zero, are given in Table 1. Although the values for Series 5 and Series 12 reach significance at 5%, the overall picture is reassuring. In no case was there an obvious pattern in the autocorrelations. Additional moving average terms were fitted to some of the series and found not to be significant.

It was particularly encouraging that the model IMA(0,1,1) appeared to fit the series from both mills, with similar values for the parameter θ. The only obvious difference between the two mills was a consistently higher value for σ_a in the series from the second mill.

3.2 SYSTEM DYNAMICS

The response of flour protein content at the sampling point to a change in the setting of the gluten feeder is delayed for two reasons. Firstly, the feeder itself does not respond instantaneously to such a change; its own internal controller adjusts the addition level gradually over a short period to match the set point. Secondly, there is the obvious delay because the sampling point is downstream of the addition point.

The dynamics of the system were investigated experimentally by making a number of large step changes to the gluten addition settings and using the NIROS instrument to observe their effect. Making a change immediately after a measurement had been taken, which is when the control system makes its adjustments, it was found that no effect was observed

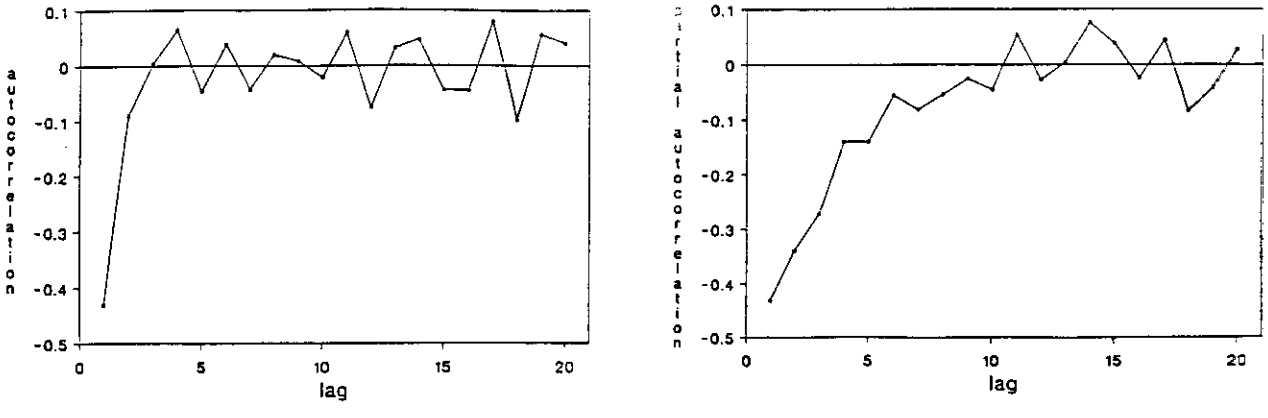

Figure 3. *Sample autocorrelation function and partial autocorrelation function for the first difference of Series 2. The approximate standard error of a sample correlation is* $1/\sqrt{766} = 0.04$.

one minute later in the next flour sample, but that the full expected effect was seen in the sample after that. Thus, the response of the system may be modelled as a simple delay of one time unit.

3.3 OPTIMAL CONTROL

Given the IMA(0,1,1) model and the simple system dynamics it is easy to derive the control strategy that minimizes mean square deviation from target. After observing z_t, an adjustment should be made to correct for a proportion $1 - \theta$ of the error (z_t – target + last adjustment). The last adjustment is involved because it will not have affected z_t although it will affect z_{t+1} and subsequent measurements.

It may seem to be a problem that the "optimal" control strategy depends on a parameter θ that varies significantly between the twelve series analyzed. Fortunately, however, the efficiency of the control system is not particularly sensitive to an incorrect specification of θ. Suppose that the control system assumes $\theta = \theta_0$ when the true value is $\theta = \theta_t$. Then the variance of the controlled output can be shown to be

$$v = \{1 + (\theta_t - \theta_0)^2 / (1 - \theta_0^2)\}\sigma_a^2.$$

With $\theta_0 = 0.75$, $v/\sigma_a^2 = 1$ when $\theta_t = 0.75$, i.e. when θ_0 is correct, but only increases to 1.05 for $\theta_t = 0.6$ or 0.9.

It was convenient to choose $\theta = 0.75$ for the control system because division by four, to calculate the adjustment, was easy to program in assembly language.

4. ASSESSMENT OF CONTROL SYSTEM

As part of an extensive assessment of the performance of the system a number of series of NIROS protein measurements were collected with the control system in operation. Root mean square deviations from target were calculated for fourteen series with lengths between 43 and 263 minutes. They ranged from 0.09% protein to 0.15%, averaging 0.12%. Since the theoretical value is σ_a the system actually performed better than would be expected from the results in Table 1. The Series 6–12 were, however, collected in the course of a few days and may have been affected by some common extra source of variability. The assessment series on the other hand were collected at intervals over a period of several months.

Under optimal control the output should in theory be the uncorrelated series a_t. The sample autocorrelation function for one of the assessment series, of length 119 minutes, is shown in Figure 4. The modified Box–Pierce statistic for the first twenty autocorrelations of this series gave a chi-square of 27.0 on twenty degrees of freedom.

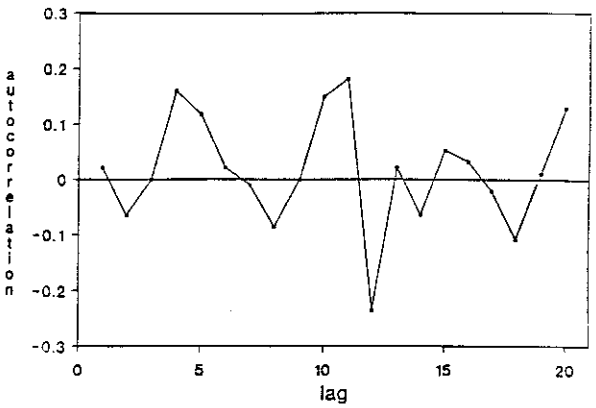

Figure 4. *Sample autocorrelation function of a series of 119 protein measurements with the process under automatic control. The approximate standard error of a sample correlation is* $1/\sqrt{119} = 0.09$.

Examination of chart recordings of the gluten addition levels during controlled production revealed considerable variation in the amount of control action needed. During some runs the addition level varied very little for long periods; in other runs quite large changes in addition levels were needed to keep the flour protein on target. Figure 5 shows a long run of the latter type. As might be expected the variation seen in Figure 5 is similar to that seen in protein levels in uncontrolled runs.

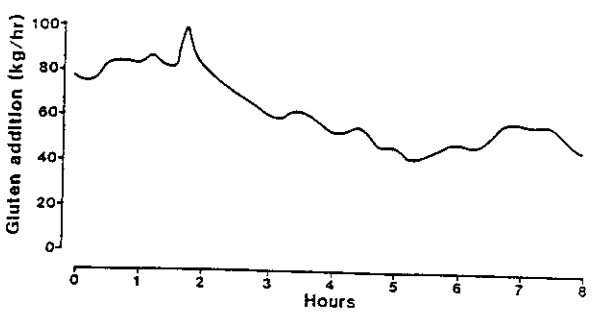

Figure 5. *Variation in the rate of gluten addition during a long production run under automatic control with a constant target.*

The overall conclusion of the assessment exercise was that the system as a whole worked well. It was taken up by a milling engineering company and is now available commercially.

5. DISCUSSION

5.1 INTERPRETING THE MODEL

Probably the most helpful way of interpreting the IMA(0,1,1) model, in this situation at least, is to note that for $\theta > 0$ it is equivalent to a Dynamic Linear Model (Harrison and Stevens, 1976) in which a system with level μ_t which evolves as a random walk

$$\mu_t = \mu_{t-1} + \delta_t$$

is observed with error

$$z_t = \mu_t + \varepsilon_t \ .$$

If the observation errors ε_t and the steps of the random walk δ_t are independently and normally distributed with zero means and standard deviation σ_ε, σ_δ respectively the correspondence with the parameters of the IMA model is $\sigma_\varepsilon = \sigma_a\sqrt{\theta}$ and $\sigma_\delta = (1 - \theta)\sigma_a$. Taking $\theta = 0.75$ and $\sigma_a = 0.12$ gives $\sigma_\varepsilon = 0.10$, which is a plausible standard deviation for the combined effect of sampling error and measurement repeatability, and $\sigma_\delta = 0.03$ for the standard deviation of the steps in the random walk.

5.2 ROBUSTNESS OF THE FEEDBACK CONTROL

Converting a correction expressed in percent protein to an adjustment of the gluten feeder requires assumptions about the protein content of the dried gluten and the flow rate of the flour stream. In practice these are likely to vary slightly for different production runs. However, the effect of misspecifying these quantities by, say, 10% is the same as that of using a value for θ that is 10% in error: a very small loss in efficiency. For the same reason the system would also be reasonably robust to an error in calibration of the gluten feeder itself, as noted in Section 2.

ACKNOWLEDGEMENTS

This research was carried out while the first author was employed by the FMBRA. Many FMBRA staff contributed to the development and testing of the NIROS gluten control system. The prototype system was installed in William King flour mill at Uxbridge, whose generous collaboration is gratefully acknowledged, as is that of the other mill involved, Heygates Ltd. at Tring. The time series work that led to the control algorithm was supported by the Ministry of Agriculture, Fisheries and Food.

REFERENCES

Box, G. E. P. and Jenkins, G. M. (1976), *Time Series Analysis: Forecasting and Control*, 2nd edn. San Francisco: Holden–Day.

Harrison, P. J. and Stevens, C. F. (1976), "Bayesian forecasting (with discussion)," *J. R. Statist. Soc. B*, 38, 205–247.

Ljung, G. M. and Box, G. E. P. (1978), "On a measure of lack of fit in time series models," *Biometrika*, 65, 297–303.