International Journal of
***O*perations *R*esearch**

International Journal of Operations Research Vol. 16, No. 4, 91-104 (2019)

Stochastic Optimization for Controlling the End State of Hydrogenation of Edible Oil in the Presence of Process Failure

Soham Chakraborty[1*], Pathik Mandal[2]

[1]SQC & OR Unit Indian Statistical Institute

203 B. T. Road Kolkata 700 108

Received February 2019; Revised November 2019; Accepted November 2019

Abstract: Controlling the end state of a growth process by terminating it at the right time is defined as a problem of passive regulation. Within this class, the periodic sampling based methods are used in many practical situations. However, it appears that this type of passive regulation system has received no systematic attention in the past. The problem of controlling the final melting point of a batch of edible oil during hydrogenation belongs to this category. This problem is formulated in terms of two stochastic optimization problems. The first problem is solved following a scenario tree based approach, while the second one is formulated as a Stochastic Dynamic Program (SDP). The SDP is solved by evaluating a quality cost function through simulation. It is shown that the performance of the hydrogenation process is expected to improve significantly under the proposed passive regulation scheme.
Keyword — Simulation, Hydrogenation of edible oil, Growth model, Control chart, Approximate dynamic programming

1. INTRODUCTION

Hydrogenation of edible oil is a process of adding hydrogen atoms by breaking the C = C double bonds that are present in the mono and polyunsaturated fatty acids. The breakage of the double bonds leads to an increase in the melting point of the oil. Thus, during hydrogenation, the melting point of a batch of oil rises gradually over time. The specifications for final melting point, which is the most important quality characteristic that needs to be controlled, is 40.0 ± 0.50°C. This is achieved by terminating the growth process at an appropriate time.

However, the hydrogenation process also suffers from occasional growth failure, which leads to the occurrence of what is known as a 'dead batch'. It is obvious that such failures need to be detected as quickly as possible so that the batch can be recovered at the earliest to keep the cost of production under control.

In order to perform the twin tasks of detection of a dead batch and controlling the final melting point, samples are drawn periodically from a batch and the melting point of the samples are determined off-line to judge the status of the batch. This gives us the data for constructing the growth curve of each batch. Figure 1 shows two typical growth curves, which also illustrate the important features of the process. For example, it shows that the batch process under consideration is a two-stage process – active hydrogenation followed by cooling. The melting point of a batch also rises slightly during cooling, but this growth cannot be controlled.

Thus, the three important decision variables for the control system are (i) The sampling points $\{t_i\}$, (ii) A variable representing the status of the batch (whether dead or active) and (iii) The cooling limit, i.e., the melting point at which the active hydrogenation phase should be terminated. Here, the control of the end state of a growth process by terminating it at the right time is called 'passive regulation'. Such a regulatory action is called passive since it does not have any impact on the growth path of the process. On the other hand, if the regulatory action is meant to influence the process path then it will be called 'active regulation'.

An important constraint of the system is the maximum number of samples that can be drawn from each batch ($nmax$ = 6). This constraint represents the capacity constraint of the testing laboratory and also the hazard associated with frequent sampling. However, it is obvious that if the cooling limit is to be used as the basis for terminating the process then the constraint on $nmax$ cannot always be satisfied. One of the two needs to be considered as a soft constraint. Here, it is decided to treat $nmax$ as the soft constraint. Another important constraint is the presence of a significant

*Corresponding author's e-mail: soham.ami@gmail.com

http://doi.org/10.6886/IJOR.201912_16(4).0001

lag in measurement (= 20 minutes).
In the existing system the process engineer decides on the most appropriate course of actions for controlling the process based on his or her subjective judgment. As a result, the performance of the process with respect to both final melting point and cycle time is found to be very poor. It is seen from Figure 2 that the distribution of final melting point is positively skewed and the melting point exceeds the upper specification limit (USL) on many occasions. The cycle time (till cooling) is also found to vary widely from 2.5 – 10 hours. The primary objective of this study is to develop an economic integrated system for on-line control of the process of hydrogenation.
For the sake of simplicity, the above control problem is split into the following two sub problems: (i) A single stage stochastic optimization problem for finding the cooling limit and (ii) A Stochastic Dynamic Program (SDP) for finding the sampling points ti. The SDP is solved following a simulation based optimization scheme. To the best of our knowledge, no previous work on this particular problem has been reported so far. The rest of the article is structured

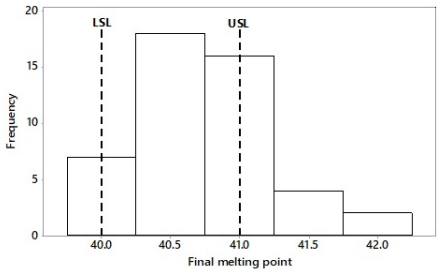

Figure 1: Two typical growth curves

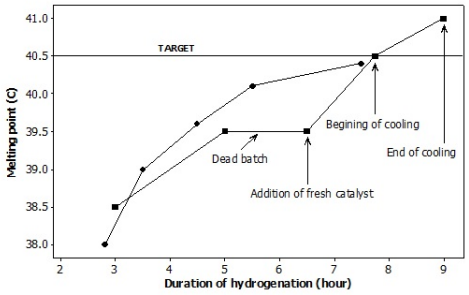

Figure 2: Histogram of final melting point (°C) along with its specification limits

as follows. Section 2 presents a literature survey. Section 3 gives a brief summary of our approach along with a description of the data. Section 4 is devoted mainly towards the mathematical formulation of the problem. Each of the steps mentioned in Section 3.1 are discussed in detail in this section. The results of our approach are presented in Section 5. The article ends with some concluding remarks in Section 6.

2. RELATED WORK

In principle, a growth process can be actively regulated to track a pre-specified path. However, in practice, such an ideal path is usually not specified. This is primarily due to the difficulty of translating the optimal performance of a growth process to a particular growth path. Nevertheless, attempts have been made in the past to specify the desired path first and then regulate the process to track the specified path (Aerts et al. (2003); Niu and Yang (n.d.)).
On the other hand, if the end time is flexible then the end state may be controlled by terminating the process at the right time, i.e., through passive regulation. Depending on the number of measurements (n) that can be made in each cycle of the process (excluding the initial state), the passive regulation problems can broadly be classified into three groups: (i) $n = 1$, (ii) $n = \infty$ (representing continuous measurement) and (iii) $1 < n \leq nmax$.
If $n = 1$, then one may use the observations on $(Y_f, X, \Delta t)$, where Y_f is the end state corresponding to the cycle time Δt and X is the vector of relevant process variables to develop a suitable model for predicting $\Delta t \mid (X, T)$, where T is the target for Y_f. The current and the past observations can then be used to develop a suitable scheme for updating the model. This is a low cost approach and is used, for example, in electroplating for controlling the coating thickness (Antony (2014); Wahab, Noordin, Izman, and Kuriniawan (2013)).
The second case represents situations where continuous monitoring of the process is possible using on-line sensors, e.g.

in-situ real time measurements for detection of the end point of a plasma etching process. Depending on the principle used to design the sensor, the amount of noise present in the data and the criticality of application, the detection of the end point may be a simple graphical exercise (Sutomo, Wang, Bullen, Braden, and Liu (2003)) or it may involve the use of sophisticated data analytic tools like Principal Component Analysis and Independent Component Analysis (Pug (2013)).
The control problem addressed in this study belongs to the last category, where $1 < n \leq nmax$. Such control problems are encountered widely in industry, agriculture and also in everyday life. For example, farmers decide on the date of harvesting based on periodic inspection. In manual cooking of rice, periodic sampling is used to decide the end point of cooking. In LD converter based steel making, samples are drawn a few minutes before the end of the blowing cycle to control the carbon content and the temperature of the bath (Mishra, Wadhwa, and Hattum (2016)). However, the results of these measurements are used for the purpose of path correction, i.e., for active regulation of the process. No published literature on passive regulation of the process based on these measurements has come to our notice.
So far as the problem of monitoring a growth process is concerned, the method of monitoring will depend on the scope of control, i.e., whether the objective is to monitor the within cycle or between cycle variation. Since a growth curve can be considered as a special type of (process) profile, the methods of profile monitoring may be used for monitoring the between cycle variation. However, in order to use these methods, it is necessary to have data in a suitable format. Presently, profile monitoring using control charts is an active area of research and numerous papers have been published on this subject (e.g. Woodall, Spitzner, Montgomery, and Gupta (2004); Mahmoud and Woodall (2004); Vaghefi, Tajbakhsh, and Norrosona (2009); Park et al. (2012)). But the number of published literature on monitoring within cycle variation is very limited (Chakraborty and Mandal (2015); Shore, Benson-Karhi, Malamud, and Bashiri (2014); Zhu et al. (2014)).

3. APPROACH AND DATA

3.1 Approach

The literature review as above suggests that a growth process can be subjected to either active or passive regulation. However, the approach of active regulation is not considered for the present process due to the large variation in its growth behavior (see Figure 3). Of course, a part of this variation can be attributed to the variation of initial melting point of the oil, but its contribution cannot be examined due to the non-availability of relevant data on initial melting point. Having decided to continue with the existing practice of passive regulation, the following approach is adopted to determine the three parameters of the control system.

(1) :Development of the growth model for the active hydrogenation phase.

(2) :Construction of the control chart based on the growth model.

(3) :Development of the model for the amount of growth during cooling.

(4) :Development of the model for growth failure.

(5) :Determination of the first sampling point (t_1) based on the failure model and specifying a distribution of melting point at t_1 (note that the melting point of a batch of oil at time $t = 0$ is unknown).

(6) :Estimation of the cooling limits based on the growth model and the power of detection of a dead batch by the control chart.

(7) :Determination of the other sampling points using all the information as above.

Strictly speaking, the control limits of the chart, the cooling limits as well as the sampling points are to be determined together for optimal performance. However, the advantage of the above simple structure is that the control chart can be used as an independent tool. The cooling limits estimated separately to simplify the solution procedure. Thus, the main optimization problem is related to the last step, which is formulated as a Stochastic Dynamic Program (SDP) and a simulation based optimization method is used to find the optimal solution.

3.2 Growth data

The growth data that is collected from the past Quality Control records of the process are observations on (MP_{ij}, t_{ij}), where MP_{ij} is the melting point of the j^th sample of the i^th batch (collected at time t_{ij}), $i = 1, 2, \cdots n(= 50)$, $j = 1, 2, \cdots n_i(\leq 6)$. Note that t_{kj} is not necessarily the same as t_{lj} for $k \neq l$. Thus, the data that is available is

scanty and is in the form of a highly unbalanced panel data. The growth of melting point for these fifty batches till the beginning of cooling is shown in Figure 3.
In addition, the time (t_0) at which fresh catalyst is added for the first time to a batch is also noted.

4. METHODOLOGY

The complete methodology is discussed under three subsections. The first subsection describes the main elements of the system. This is followed by a description of various aspects related to the formulation of the problem and the last subsection gives the solution methodology.

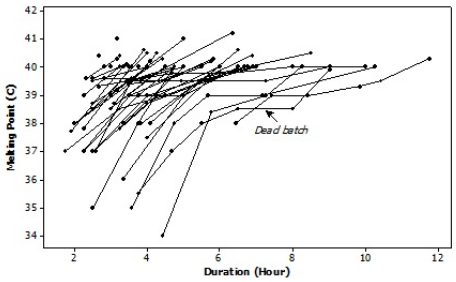

Figure 3: Growth curves of fifty sample batches till the beginning of cooling

4.1 Main elements of the system

The main elements of the system that are described in this section are the growth model, the failure model, the control chart for detection of dead batches and the distribution specifying the initial condition.

4.1.1 Process model for the active hydrogenation phase

In an earlier work, Chakraborty and Mandal (2015) have shown that the growth of melting point (MP) during hydrogenation can be described by the first order model given by

$$\mathrm{d}(MP)/\mathrm{d}t = k(MP_{eq} - MP) \tag{1}$$

where k is the growth coefficient and MP_{eq} is the equilibrium melting point. Approximating (1) by the forward difference equation, the process may be modeled as a random model with variable drift

$$MP_{t+\Delta t} = MP_t + G|(MP_t, \Delta t) \tag{2}$$

where the growth G during a time interval Δt is the variable drift component and is given by

$$G = \mathrm{d}_0(\Delta t) - \mathrm{d}_1(MP_t)(\Delta t) \tag{3}$$

Now, using the growth data as above, it is a straightforward task to estimate the parameters of (3) following the method of multiple regression. However, it is to be noted that the estimates of growths thus obtained will be biased since $G|(MP_t, \Delta t) > G|(MP_t, \Delta t) + G|(MP_{t+\Delta t_1}, \Delta t_2), \Delta t = \Delta t_1 + \Delta t_2$. In other words, since the value of Δt will vary in practice, the estimated growth path becomes dependent on the sampling scheme. Clearly, this is undesirable. Accordingly, the amount of growth, i.e., $G|(MP_t, \Delta t)$ is estimated following an innovative method of instrumental variable estimation. The details of this method are given in Chakraborty and Mandal (2018). Here we present only the final model in a compact form:

$$\begin{aligned}
&E(G^I) = 0.0805 + 0.9089G_1 \\
&G_1 = -0.0482 + 1.21819I \\
&I = \sum_{i=1}^{k} g_i, g_i = 0.852(\Delta t) - 0.4563(MPS)(\Delta t), MPS = (MP - 38.488)/1.082 \\
&\Delta t = \sum_{i=1}^{k} \Delta t_i, \Delta t_{i<k} = 0.5, \Delta t_k = (\Delta t - 0.5(k-1)) < 0.5 \\
&SN(G^I) = 113.091 - 2.85(MP) + 2.57(t - t_0) = 10\log(E(G^I)/\sigma(G^I))^2
\end{aligned} \tag{4}$$

where MPS is the standardized value of MP, t_0 is the time of first catalyst addition and I is the instrumental variable that is constructed from the observed growth data. The mean, signal-to-noise ratio and standard deviation of $G|(MPt, \Delta t)$ are represented by $E(G^I)$, $SN(G^I)$ and $\sigma(G^I)$ respectively. It is to be noted that the above model is developed using growth data of only genuine active intervals, i.e., the intervals which are followed by another interval of positive growth without addition of fresh catalyst. Also, for the sake of accuracy, the model will be used for estimating growth within an interval of $0.5 \leq \Delta t \leq 1.5$hour.

4.1.2 Control chart for detection of dead batches

The control chart is developed based on the process model given by (4), which defines the in-control state of the process. Next, an individual control chart is constructed based on the control statistic

$$Q = (G - \hat{G^I})/\hat{\sigma} \tag{5}$$

where the estimates $\hat{G^I}$ and $\hat{\sigma}$ are obtained from the process model (4). The other details of the chart including its power of detecting dead batches are given elsewhere (Chakraborty and Mandal (2018)). Here we only note that the center line of the chart is zero and the lower and upper control limits are placed at -2.35 and 2.6 respectively. If a point falls below the lower control limit, then it is considered as a signal for a dead batch and fresh catalyst needs to be added to the batch for its recovery. On the other hand, if a point falls above the upper control limit then the process engineer has to make his or her own subjective judgment for regulating the process.
The above control chart has good power of detecting a dead batch with $\Delta t = 1$ hour. It is also found that the power of detection becomes negligible with $\Delta t \leq 30$minutes and the detection becomes almost certain with $\Delta t \geq 75$ minutes.

4.1.3 Distribution of growth during cooling

Let the amount of growth during cooling be denoted by y' and the cooling decision be based on the sample drawn at time t' Since there is a measurement lag of twenty minutes, we have

$$MP_f = MP_{t'} + \Delta MP_{lag} + y' = MP_{t'+lag} + y' \tag{6}$$

where MP_f is the final melting point of the batch and $lag = 1/3$ hour. Now, assuming that a batch is active throughout the growth process, the estimates of $MP_{t'+lag}|(MP_{t'}, \Delta t = lag)$ are obtained from the process model (2), where G is given by (4). These estimates are then substituted in (6) to obtain the estimates of y'. However, it is observed that the growth during cooling is zero in as many as six out of the forty seven batches considered for developing the process model (4). Excluding these six batches, the estimates of $y' > 0$ gives a very good fit to the two-parameter Weibull distribution (Figure 4). Thus, the following mixture of Bernoulli and Weibull is specified as the distribution of y',

$$f(y') = \begin{cases} p', y' = 0 \\ (1-p')g(y'), y' > 0 \end{cases} \tag{7}$$

where p' is the probability of no growth during cooling (= 6/47) and f and g are the probability density functions. The density $g(y')$ of the Weibull distribution is given by

$$g(y') = (\beta/\alpha)(y'/\alpha)^{\beta-1} \exp[-(y'/\alpha)^{\beta}], \alpha > 0, \beta > 0, y' > 0 \tag{8}$$

and its mean and variance are given by

$$\mu' = \alpha\Gamma(1 + 1/\beta), \sigma^2[\Gamma(1 + 2/\beta) - \Gamma^2(1 + 1/\beta)] \tag{9}$$

Now, since the estimate of the shape parameter $(\hat{\beta})$ is 1.504 and that of the scale parameter $(\hat{\alpha})$ is 0.588 (see Figure 4), we have $(\hat{\mu'})$ and $(\hat{\sigma'^2})$ Further, it can be easily shown that

$$E(y') = (1-p)\mu' \text{ and } \text{Var}(y') = (1-p')(p'\mu'^2 + \sigma'^2) \tag{10}$$

which gives the estimates of mean and variance of y' as 0.4628 and 0.144 respectively. These estimates are used later for finding the cooling limits.

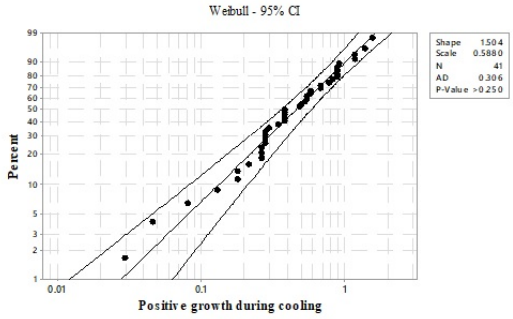

Figure 4: Weibull probability plot of $y' > 0$

4.1.4 Distribution of failure time

It is found from the control chart that seven out of the fifty sample batches can be considered as dead. Thus, the failure data consists of seven interval-censored and forty right-censored observations. It is also noted that the failure time (t_d) and the time of first catalyst addition t_0 are exceptionally high for a few batches. This suggests that the process has remained idle for some time, perhaps due to some process problems like malfunctioning of the vacuum pump. Consequently, instead of t_d , it is decided to model the distribution of $(t_d - t_0)$. The methods of estimating the parameters of a life distribution based on censored data are available in many books on reliability (e.g. Meeker and Escobar, 1998). Here the commercial software MINITAB is used for developing the failure model. It is seen from Figure 5 that the Log-normal distribution gives a good fit to the observations on $(t_d - t_0)$. The least square estimates of the location and scale parameters of the distribution are also shown in this figure.

4.1.5 Initial condition

The initial melting point of a batch is neither measured nor known. The earliest a sample has been drawn in the fifty sample batches is at $t = 1.75$ hour. However, it is important to detect the status of a batch as early as possible for controlling the process efficiently. The failure distribution developed in the previous section suggests that the probability of a batch failure before 1.5 hour is nearly zero. Therefore, it is decided to take the first sample at $t = 1.5$ hour. Next, the distribution of melting point at $t = 1.5$ hour is obtained as follows. Let, for a given batch, the melting

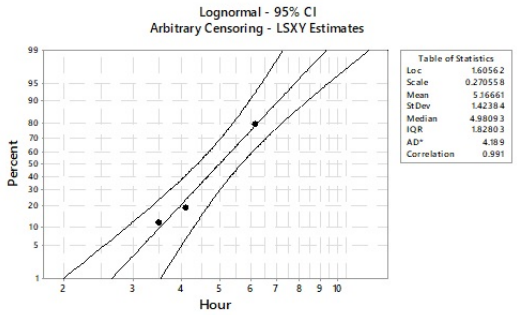

Figure 5: Lognormal probability plot of failure time $(t_d - t_0)$

point of the first sample drawn at $t_1 > 1.5$ be denoted by MP_1. Then the desired estimate MP_0 is obtained by solving $E(MP|(MP_0, \Delta t = t_1 - 1.5)) = MP_1$ Also, considering the risk of extrapolation and the fact that the estimates of growth obtained from (4) will be poor with $\Delta t > 2$ hours, the backward extrapolation is limited only to those batches for which $t_1 \leq 3.5$. The distribution of the estimates thus obtained is shown in Figure 6. It is seen that the distribution is negatively skewed. Although the sample size here is very small, it appears that a stable distribution at $t = 1.5$ hour may not exist. But, for developing the control scheme, it is necessary to specify a reasonable initial distribution that will be valid under a wide range of operating conditions. Also, for the purpose of simulation, it is preferable to have a bounded distribution. Considering the above, the following truncated normal distribution is specified as the distribution of $MP_{t=1.5}$,

$$MP_{t=1.5} \bigcap TN(36.8, 2, 32, 39.6) \tag{11}$$

where the parameters signify that a normal distribution with mean 36.8 and standard deviation 2 is truncated at 32 and 39.6. The mean and standard deviation of the above distribution are very close to the mean and standard deviation of the observed distribution (Figure 6).

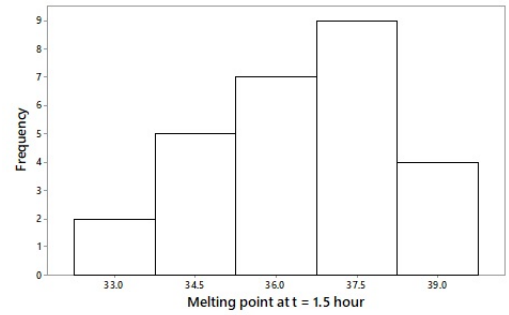

Figure 6: Histogram of melting point at $t_1 = 1.5$ hour

4.2 Problem formulation

4.2.1 State variables

Let S_0 represents the true status of a batch being hydrogenated, i.e., $S_0 = Active, Dead$. The observed status of the batch is defined similarly and is denoted by S_1. However, since the observation on S_1 is made by a control chart, it will be unrealistic to neglect the error involved in this measurement. So let the four possible scenarios related to this measurement be denoted by $S = s_1, s_2, s_3, s_4$, where $s_1 = Deadanddetected$, $s_2 = Deadbutundetected$, $s_3 = Activeandnotidentifiedasdead$, and $s_4 = Activebutidentifiedasdead$. The quality cost associated with a particular growth path is evaluated taking all these four possible scenarios into account.
The measurement error associated with MP is included in the variance of its estimate.

4.2.2 Cost function

Let the total quality cost (C) associated with a growth path (ω) be given by

$$C(\omega) = C_P(\omega) + C_E(\omega) \tag{12}$$

where C_P is the path cost and C_E is the end cost. The path cost C_P has two components,

$$C_P = C_1 + C_2 \tag{13}$$

where C_1 is the cost of sampling and testing and C_2 is the failure cost. Further, the failure cost also has two components - detection cost (C_d), i.e., the cost of delay in detecting a dead batch and the cost of recovering (C_r) a batch that has been identified as dead, whether rightly or wrongly. Thus, we have

$$C_2 = C_d + C_r \tag{14}$$

Now, let a batch that becomes dead at time t_d is detected at the j^th sample. Since there is a measurement lag of twenty minutes, the delay in detection or the failure time (FT) is given by $FT = t_j + lag - t_d$. It is possible that a batch may fail between $(t_j, t_j + lag)$ after it is decided to cool the batch based on the sample drawn at t_j. There is a cost (in the form of loss of information) associated with such failures too but such costs (being very small) are not taken into account.
The cost of recovery is assumed to be proportional to RT, the time lost in making arrangement for addition of fresh catalyst. It is estimated that RT is approximately 10 minutes per failure. The cost of catalyst is not considered since the catalyst remains in the system (recycled). Further, it is assumed that a batch can become dead at most once before cooling.
The end cost is the quality loss component and is assumed to be given by the quadratic loss function

$$C_E = \lambda(MP_f - 40.5)^2 \tag{15}$$

where λ is the loss coefficient and 40.5 is the target for MP_f. It is to be noted that if a dead batch is not recovered before cooling, then MP_f is the same as the melting point at t_d, otherwise it is given by (6).
The total cost of quality for a batch is then given by

$$C = _{C_1}(NS + 1) + _{C_2}(FT + RT) + \lambda(MP_f - 40.5)^2 \tag{16}$$

where c_1 is the cost of sampling and testing per sample, NS is the number of samples drawn till cooling and c_2 is the hourly cost of downtime. The economics of the process suggests that c_2 is approximately 100 times of c_1 and λ is approximately 3.5 times of c_2. Thus, the cost function becomes

$$C(\omega) = (NS + 1) + \sum_{j=1}^{NS(\omega)} [100(a_j l_j + b_j(RT))] + 350(MP_f - 40.5)^2 \tag{17}$$

where t_j, j = 1, 2, $\cdots$, NS are the time points at which the samples are drawn till cooling and $a_j = 0$, if the batch is active at t_j and 1, otherwise. The failure time l_j for the j^th interval (t_{j-1}, t_j) is given by

$$l_j = \begin{cases} t_j + lag - t_d, & \text{if } t_{j-1} < t_d \leq t_j \text{ [i.e., } S_j = s_1 \text{ or } s_2] \\ t_j - t_{j-1}, & \text{if } t_d < t_{j-1} \text{ [i.e., } S_{j-1} = s_2] \end{cases} \tag{18}$$

Also, $b_j = 1$ if $S_j = s_1$ or s_4 and $MP_j < MP_C$, where MP_C is the cooling limit.

4.2.3 Penalty function

It has already been noted in Section 1 that the constraints on both MP_C and $nmax$ cannot be satisfied simultaneously. There will be instances of MP_j being less than MP_C at $j = nmax$. So it is decided to treat the limit on $nmax$ (= 6) as a soft constraint, where $nmax$ includes the last sample taken at the end of the cooling phase. Thus, for the purpose of optimization, the following penalty function is added to the original cost function [Equation (17)],

$$C_p(\omega) = \begin{cases} 100(NS(\omega) - 5)^3, & \text{if } NS > 5 \\ 0, & \text{otherwise} \end{cases} \tag{19}$$

4.2.4 Mathematical formulation of the problem

Let, for a given batch, the sampling intervals are defined by $\Delta t_j = t_j - t_{j-1}, j = 2, 3, \cdots, NS$. Also let A, D and R be the set of all active, dead and recovered intervals respectively, i.e., A contains the intervals in which no failure occurs, D contains those intervals in which the batch fails at time $t_{j-1} \leq t_d \leq t_j$, and R is the set of intervals in which the control chart gives an out-of-control signal. For a given batch, the set D is either empty or has at most one element (by assumption). Further, the variable t_0 is generalized to t_{0j} to consider the time of fresh catalyst addition for each interval. The SDP can now be formally stated as follows:

Minimize $E[C(\omega) + C_p(\omega)]$
subject to

(i) $t_i = 1.5$.

(ii) The distribution of MP at t_1 is given by (13).

(iii) At t_1, S_0 = Active.

(iv) $MP_j = MP_{j-1} + G_j, \quad j = 2, 3, \cdots, NS,$
$G_j \bigcap N(\mu_j, \sigma_j)$, where μ_j and σ_j are given by (6) with
$\Delta t = t_j - t_{j-1}, \quad j \in A,$
$\quad = t_d - t_{j-1}, \quad j \in D,$
$t_{0j} = 0, \quad j - 1 \in R,$
$\quad = 1, \quad$ otherwise.

(v) The distribution of $t_d - t_0$ is given in Figure 6.

(vi) The mapping $S_{0j} \rightarrow S_1, \quad j = 2, 3, \cdots, NS$ is obtained by the control chart.

(vii) $MP_{NS} > MP_C$, where the cooling limits MP_C are to be specified.

(viii) $MP_f = MP_{t'+lag} + y'$ where the distribution of y' is given by (9).

(ix) lag = 1/3

(x) $0.5 \leq \Delta t_j \leq 1.5; \quad \Delta 0.5$ if $j - 1 \in R$

It may be noted from the last constraint that the length of the sampling intervals post recovery has been limited to thirty minutes. This is to have some safeguard against the possibility of greater variation in growth behaviour of the process when fresh catalyst is added for the second time, particularly when such additions are made unnecessarily. In other words, the controller operates in the safe mode post recovery. This is similar to the control schemes proposed by Mozumder, Saxena, and Collins (1994) and Sachs, Hu, and Ingolfsson (1995), where the control chart is used for selecting an appropriate mode of drift control

4.3 Solution methodology

In principle, the SDP can be solved to determine the cooling limits and the sampling intervals simultaneously. However, in order to simplify the solution procedure, the cooling limits are determined separately. Such an approach has an added advantage. If the cooling limits are found to be too low (say < 39.50°C), then it will be necessary to put a constraint on MP_C and then solve the SDP. This is because if a batch becomes dead immediately after cooling then the end cost cannot be evaluated properly by the quadratic loss function in such cases. The losses are expected to be much higher than that given by the loss function. An independent exercise to find the cooling limits will then be helpful in specifying a suitable constraint on MP_C, if needed.
However, in the present case, the standalone cooling limits obtained from an independent exercise are judged to be satisfactory. These limits are then used for finding the optimal sampling intervals.

4.3.1 Optimal cooling limits

Let the present state of the process be denoted by MP_t. Then the determination of cooling limit involves a comparison of cost of cooling a batch at the present state with that of going ahead by one more step. The minimum permissible step size is chosen as half an hour so that the limits remain close to the target. Thus, the determination of cooling limit becomes a single stage stochastic optimization problem with only two alternatives - start cooling at MP_t or continue hydrogenation for another thirty minutes. Of course, the costs will vary depending on whether the batch, at time t, is identified as active or dead. If it is identified as dead, then the options are either to start cooling at MP_t without recovering the batch or to recover it and then continue hydrogenation for thirty minutes.
The scenario tree for the case when the batch at (MP, t) is identified as active by the control chart is shown in Figure 7. The two simplifying assumptions made in constructing the scenario tree are as follows: (i) The control chart cannot detect a dead batch if the sample is drawn less than half an hour after the occurrence of death and (ii) Second sample detection of growth failure is certain. It is already noted in the section describing the control chart that these two assumptions are very reasonable. Thus, the expected total cost of cooling at MP_t is given by (see Figure 7)

$$E(TC_1) = (1 - p_1)C_A + p_1C_D$$

where p_1 , by virtue of assumption (i) above, is approximately the probability of a batch becoming dead in the interval (t–0.5, $t + lag$), given that it is active at t–0.5.
On the other hand, instead of cooling at $(t + lag)$ if it is decided to continue hydrogenation for another thirty minutes then the expected cost is given by (see Figure 7)

$$E(TC_2) = (1 - p_1)[p_2C_{AD} + (1 - p_2)C_{AA}] + p_1C_{DA}$$

where p_2 is the probability of failure in the interval $(t + lag, t + lag + 0.5)$, given that the batch is active at $(t + lag)$.
The cost variables $(C_A, C_D, \cdots, C_{DA})$ involved in the above two equations are described in Table 1.
To illustrate, let us consider the cost component C_A, which is given by

$$\begin{aligned} C_A &= C_1 + C_d + C_r + C_E = 350[E(MP_f - 40.5)^2)] \\ &= 350[E(MP_{f+lag}) + E(y' - 40.5)^2 + V(MP_{t+lag}) + V(y')] \end{aligned}$$

where $E(MP_{t+lag})$ and $V(MP_{t+lag})$ are obtained from (6) and $E(y')$ and $V(y')$ are obtained from (10). The other costs are evaluated similarly.
The desired optimal cooling limit for a given t is given by the minimum melting point at which $E(TC_2)$ becomes greater than $E(TC_1)$.
The scenario tree for the case of a dead batch, i.e., when $S = s_1, s_4$ is constructed similarly. Here the first two branches at $(t + lag)$ correspond to the batch being genuinely dead or it is a false alarm. It may be noted that the probability of a false alarm at the lower limit of the control chart (= -2.35) is very low (< 1%). Hence this branch can be ignored without having any significant impact on the final results. However, here no such approximation is made for computing the expected costs. The method of computation remains the same as before.

Table 1: Cost variables for determining the cooling limits

Cost variable	Description	Cost component*			
		C_1	C_d	C_r	C_E
C_A	Expected cost of cooling an active batch at $(t+lag)$	0	0	0	$\alpha_2 E(MP_f - 40.5)^2$ $MP_f = MP_{t+lag} + y'$
C_D	Expected cost of cooling a dead batch at $(t+lag)$	0	0	0	$\alpha_2 E(MP_f - 40.5)^2$ $MP_f = MP_t$
C_{AA}	Expected cost of cooling an active batch after hydrogenation for 0.5 hour	1	0	0	$\alpha_2 E(MP_f - 40.5)^2$ $MP_f = MP_{t+lag+0.5} + y'$
C_{AD}	Expected cost of cooling a dead batch after hydrogenation for 0.5 hour given that it was active at the beginning	1	$0.25\alpha_1$	0	$\alpha_2 E(MP_f - 40.5)^2$ $MP_f = MP_{t+lag+0.25}$
C_{DA}	Expected cost of cooling a recovered batch after hydrogenation for 0.5 hour given that it failed $0.5 + RT$ hour before	2	$\alpha_1(0.5+lag)$	$k_1(RT)$	$\alpha_2 E(MP_f - 40.5)^2$ $MP_f = MP_{t+0.5} + y'$ $t_0 = 0$

*C_1 =Sampling cost, C_d = Detection cost, C_r = Recovery cost, C_E = End cost, $\alpha_1 = 100$, $\alpha_2 = 350$

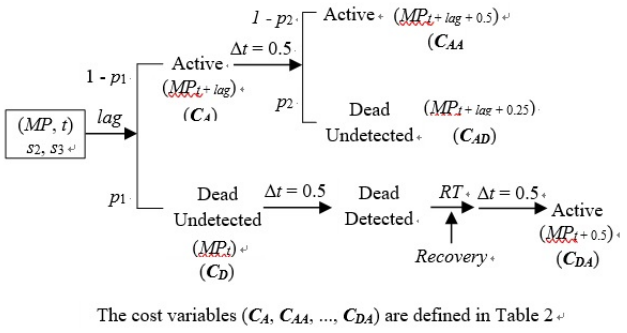

Figure 7: Scenario tree showing the possible states at the time of cooling when a batch is identified as active based on the sample drawn at time t

4.3.2 Optimal sampling intervals

It is obvious from the above formulation of the SDP that even if the cooling limits are specified beforehand, some sort of approximation needs to be made for determining the optimal sampling intervals. Therefore, the method of Approximate Dynamic Programming (ADP) is used for solving the SDP. Powell (2010) has classified the approximations comprising the body of ADP into four categories: (i) Myopic cost function approximation, (ii) Look-ahead policy (iii) Policy function approximation and (ii) Value function approximation. For example, the specification of $t_1 = 1.5$ hour can be thought of as an outcome of the myopic cost function approximation. The optimal sampling intervals are obtained following a policy function approximation described below.

Policy functions. The policies chosen for determining the sampling intervals (DT) are as follows:

$$P1 : DT = k_1(PT), \;\; 0 < k_1 \leq 1 \tag{20}$$

$$P2 : DT = k_2(PT), \;\; 0 < k_2 = b_0 + b_1(MP_t) + b_2(t - t_0), \;\; 0 < k_2 \leq 1 \tag{21}$$

where PT is the predicted time to reach the melting point of 400°C from the present state of MP_t and k_1, k_2 are shrinkage factors. The predicted time is obtained from the process model (4). However, in order to satisfy the constraint $0.5 \leq \Delta t \leq 1.5$, the final sampling intervals are obtained as

$$\begin{aligned} \Delta t &= 1.5, \;\; \text{if} \;\; DT > 1.5 \\ &= 0.5, \;\; \text{if} \;\; DT < 0.5 \\ &= DT, \;\; \text{otherwise} \end{aligned}$$

Thus the task now reduces to finding the optimal values of the parameters of the policy functions. As already noted, a simulation based optimization strategy is used for estimating the parameters of the policy functions.

Simulation algorithm. The high level diagram of the simulation process is shown in Figure 8 and a summary of the simulation algorithm is given below. The algorithm is implemented in R.

Step 1: Generate a random initial MP from (11). $t = 1.5$

Step 2: Generate a random failure time t_d

Step 3: Compute DT using (20) or (21). Use $DT = 0.5$ for a recovered batch

Step 4: Predict MP at $t = t + DT$

Step 5: Determine status of the batch using the control chart

Step 6: Go to step 8 if $MP \geq MP_C$, otherwise continue

Step 7: Recover batch, if the control chart gives a dead signal

Step 8: Compute quality cost components for the interval and go to step 3

Step 9: Generate a random y' from (7)

Step 10: Compute final melting point, the end cost and the penalty cost

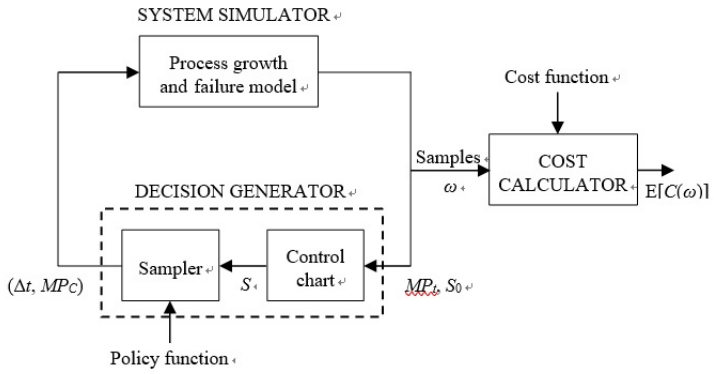

Figure 8: High level diagram of the simulation process

Optimization of the parameters of $P1$ and $P2$. The parameter k_1 of $P1$ is optimized in a straightforward manner. The average of $C(\omega)$ is computed based on 25000 samples for a series of values of k_1 with a step size of 0.05. The value of k_1, at which the average becomes minimum, is the desired solution.
However, the estimation of the parameters of $P2$ is not such a simple task. The method of Experimental Regression Analysis proposed by Taguchi (1987) is used for this purpose. In this method, the parameter space in which the optimal is expected to lie is systematically searched using a statistical design to locate the optimum. At the first iteration, the levels of the factors (parameters) cover the entire space that needs to be searched. In subsequent iterations the levels of the parameters are selected around the best levels of the previous iteration. The process ends when there is no significant improvement in response. Here, a 3^3 design is used for the purpose of experimentation. For the first iteration, the total cost is evaluated based on 2000 samples and the sample size is increased gradually to 25000 for the last (4^{th}) iteration.

5. RESULTS

5.1 Cooling limits

The cooling limits are obtained following the methodology described above. The comparison of costs for the two options when a batch is identified as active is shown in Table 2. The corresponding cooling limits are also indicated in this table. A similar table is constructed for determining the cooling limits when a batch is identified as dead. The cooling limits thus obtained are summarized in Table 3. It is seen from this table that the cooling limit increases with the increase in t. This is to be expected since the variance of melting point decreases with the increase in t (see the SN function in (4)). The cooling limit for a dead batch is the highest for obvious reason of zero growth during cooling for a dead batch.

Table 2: Comparison of costs for determining the cooling limits* (active batches)

t	Cost of cooling at MP_t					Cost of continuing hydrogenation				
	MP					MP				
	39.6	39.7	39.8	39.9	40.0	39.6	39.7	39.8	39.9	40.0
2	**78.9**	65.9	57.8	56.2	79.2	74.5	72.4	73.2	76.7	
3	76.6	**63.8**	55.9	52.9	54.8	68.9	65.3	64.4	66.3	70.9
4	87.9	72.4	**61.9**	56.4	56.0	74.3	68.9	66.4	66.7	69.9
5	121.7	98.5	80.7	**68.3**	61.2	92.0	84.2	79.3	77.3	78.2
6	153.1	122.8	98.2	**79.4**	66.2	106.5	97.5	91.3	88.1	87.9
7	172.8	138.1	109.3	**86.4**	69.4	114.9	105.4	98.9	95.2	94.6
8	184.2	147.0	115.7	**90.5**	71.2	119.6	110.0	103.2	99.4	98.6

*The melting points corresponding to the bold faced numbers are the cooling limits

Table 3: Cooling limits

Active batch				Dead batch
$1.5 \leq t \leq 2.5$	$2.5 < t \leq 3.5$	$3.5 < t \leq 4.5$	$t > 4.5$	All t
39.6	39.7	39.8	39.9	40.1

5.2 Optimal policies

Figure 9 shows the plot of $E[C(\omega)]$ versus k_1. It is obvious from this figure that the optimal value of k_1 is 0.75. The expected cost for $k_1 < 0.5$, which are much higher than that for $k_1 = 0.5$, are not shown in this figure.
The optimal policy $P2$ turns out to be simpler than specified, since the optimal value of b1 is found to be zero. The optimal shrinkage factor k_2 is given by $k_2 = 0.65 + 0.02(t–t_0)$.

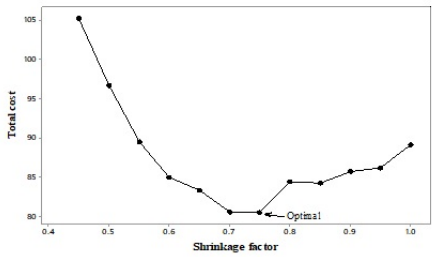

Figure 9: Optimal shrinkage factor k1 for minimum cost

5.3 Comparison of performance of P1, P2 and the existing procedure

Table 4 summarizes the performance of the two policies (optimized) and the existing procedure. It is seen that there is no significant difference in performance between $P1$ and $P2$. However, since the performance of $P2$ is found to be numerically better than $P1$ with respect to all the three important parameters, namely total cost, cycle time and % detection of dead batches, the policy $P2$ is recommended for implementation. Next, the performance of $P2$ is compared with that of the existing system (E). It is seen that the proposed policy is far superior to the existing one. A comparison of the individual cost and other elements reveals the following: (i) The similarity of the penalty cost of $P2$ and E suggests that the concern related to the workload of the testing laboratory has been addressed adequately. (ii) The sampling cost of $P2$ is marginally higher than that of E. This is to be expected since an extra sample is taken at $t = 1.5$ hour for every batch. However, the additional information provided by this extra sample about the growth behavior of the process over a larger range of t is likely to be very valuable in developing a better process model in future. (iii) Both the failure and recovery cost of the existing system is significantly higher than that of $P2$. This also indicates the effectiveness of the control chart (compared to the existing naive approach) in detecting a dead batch. (iv) Most importantly, the optimal system $P2$ is expected to reduce the quality loss by about 46% from the present level. However, even under $P2$, the quality loss remains the main contributor to the total cost (64.2%). Therefore, the future improvement effort should be directed towards reduction of this component. (v) The policy $P2$ is also expected to reduce the cycle time by more than half an hour, which is very significant from the point of view of production cost. This reduction in cycle time is a combined effect of lower failure time, less overshooting of the final melting point and recovery of the extremely slow growing batches.

Table 4: Comparison of performance of $P1$, $P2$ and the existing system

Item	Average		
	P1	P2	Existing
Total cost	80.54	79.03	128.95
Penalty cost	8.56	8.9	6.38
Sampling cost	4.34	4.34	3.6
Failure cost	15.47	14.57	20.29
Recovery cost	1.46	1.46	4.26
End cost	51.71	50.76	94.42
Mean (MP_f)	(40.49)	(40.48)	(40.71)
MSE (MP_f)	(0.148)	(0.145)	(0.27)
Cycle time*	(3.47)	(3.39)	(4.0)
% Detection	(71.76)	(75.51)	-

*From t_0 to the beginning of cooling

5.4 Look-up table

In order to facilitate the implementation of $P2$, a two-way look-up table giving the values of Δt for various combinations of MP and t is prepared. An abridged version of the same is shown in Table 5.

Table 5: Look-up table for Δt in minutes (when the batch is identified as active*)

MP	T				
	1.5	2.5	3.5	4.5	6
32-38.5			90		
38.6	87	90	90	90	90
38.8	79	82	84	86	90
39.0	70	72	74	76	79
39.2	59	61	63	65	67
39.4	54	56	58	59	61
39.6	41	42	43	44	46
39.8-39.9			30		

*Use $\Delta t = 30$ for a recovered batch

6. CONCLUSION

An integrated control system is proposed for controlling the final melting point of a batch of partially hydrogenated edible oil taking the effect of growth failures into account. The approach proposed is fairly general and can be used in many other similar situations. It can also be generalized to situations where the growth rate is slowed down deliberately towards the end of process for achieving better control of the end state.

REFERENCES

Aerts, J., Buggenhout, V., Vranken, E., Lippens, M., Buyse, J., Decuypere, E., & Berkmans, D. (2003). Active control of the growth trajectory of broiler chickens based on online animal responses. *Poultry Sc*, *82*, 1853-1862.

Antony, J. (2014). *Design of experiments for engineers and scientists, 2nd edition.* Amsterdam: Butterworth Heinemann.

Chakraborty, S., & Mandal, P. (2015). Spc based on growth models for monitoring the process of hydrogenation of edible oil. *J. Food Eng*, *146*, 192-203.

Chakraborty, S., & Mandal, P. (2018). *Control charts for detection of dead batches during hydrogenation of edible oil.* [Technical report SQCOR-2018-06]. SQC & OR Division, Kolkata..

Mahmoud, M. A., & Woodall, W. H. (2004). Phase i analysis of linear profiles with calibration applications. *Technometrics*, *46*, 380-391.

Mishra, S., Wadhwa, M., & Hattum, G. V. (2016). *Sublance: Ultimate tool for steel making.* [Technical Report].

Mozumder, P. K., Saxena, S., & Collins, D. J. (1994). A monitor wafer based controllers for semiconductor processes. *IEEE Trans. Semiconductor Manuf* , *7*, 400-411.

Niu, L., & Yang, D. (n.d.). Dynamic simulation and optimization for batch reactor control profiles. *Intl. Proc. Comp. Sc. Info. Tech. (IPCSIT)*, *22*, 58-68.

Park, M., Kim, J., Jeong, M. K., Hamouda, A. M. S., Al-Khalifa, K. N., & Elsayed, E. A. (2012). Economic cost models of integrated apc controlled spc charts. *Intl. J. Prod. Res*, *50*(14), 3936-3955.

Powell, W. B. (2010). *Approximate dynamic programming.* New York: John Wiley & Sons.

Pug, C. J. (2013). *End point detection in reactive ion etching.* (Dissertation, University College London).

Sachs, E., Hu, A., & Ingolfsson, A. (1995). Run-by-run process control: Combining spc and feedback control. *IEEE Trans. Semiconductor Manuf* , *8*, 26-43.

Shore, H., Benson-Karhi, D., Malamud, M., & Bashiri, A. (2014). Customized fetal growth modelling and monitoring – a statistical process control approach. *Qual. Eng*, *26*, 290-310.

Sutomo, W., Wang, X., Bullen, D., Braden, S. K., & Liu, C. (2003). Development of an end- point detector for parylene deposition process. *J. Microelectromech. Sys*, *12*, 64-69.

Taguchi, G. (1987). *System of experimental design, vol. 1.* New York: UNIPUB/Kraus.

Vaghefi, A., Tajbakhsh, S. D., & Norrosona, R. (2009). Phase ii monitoring of nonlinear profiles. *Comm. Stat. – Theory and Methods*, *38*, 1834-1851.

Wahab, H. A., Noordin, M. Y., Izman, S., & Kuriniawan, D. (2013). Quantitative analysis of electroplated nickel coating on hard metal. *The Scientific World Journal 2013*.

Woodall, W. H., Spitzner, D. J., Montgomery, D. C., & Gupta, S. (2004). Using control charts to monitor process and product quality profiles. *J. Qual. Tech*, *36*(3), 309-320.

Zhu, L., Dai, C., Sun, H., Li, W., Jin, R., & Wang, K. (2014). *Curve monitoring for a single crystal ingot growth process. 5th intl.* Innov., Xi'an, China: Asia Conf. Ind. Eng. Mgmt.