9th INTERNATIONAL SYMPOSIUM ON MEASUREMENT AND QUALITY CONTROL (9th ISMQC)
November 21 – 24, 2007, IIT Madras

OPTIMAL MEASUREMENT SCHEDULE TO SUPPORT QUALITY MANAGEMENT

Grén J., Konkarikoski K. and Ritala R.
Tampere University of Technology,
Institute of Measurement and Information Technology,
P.O. Box 692, FIN-33101 Tampere, Finland
kimmo.konkarikoski@tut.fi, ph +358 40 5794393

Abstract: *In industrial processes, such as in papermaking, the product quality is measured from end product samples in laboratory. It is a common practice to measure all quality parameters at regular time intervals. However, this is costly and may limit possibilities to measure those quality parameters that would be most important for overall uncertainty management and thus for decision support. This paper show general theory for measurement information dynamics and discusses how we can find the lowest cost measurement schedule satisfying the constraints on uncertainty in quality information. This paper gives theoretical approach with real data application and results show how optimal measurement schedule can be obtained to support quality management.*
Key words: *Decision support, quality, uncertainty, Bayesian, information.*

1. INTRODUCTION

The process industries make use of hundreds of on-line and laboratory measurements to monitor and control the process (Latva-Käyrä, 2003). Information systems are designed with the aim of supporting the daily decision making about the process and product quality by operators and engineers so that the best practice of operation can be achieved continuously. Measurements, soft sensors and process simulators form the basis for such decision support by reducing the uncertainty about the present state of the process and about its future evolution.

In process industries, such as papermaking, the quality management is commonly based on three level hierarchical measurement structures: accurate but costly and infrequent laboratory measurements, automated quality analyzers sampling more frequently and mimicking laboratory analyses, and indirect but frequent on-line measurements for automatic control. In paper mills, measuring frequency of analyzers is usually once per machine reel, or 1-3 times an hour, whereas laboratory analyses are made at most 3 times a day. These frequencies are to be compared with that paper web produced continuously at web speed of up to 30 m/s, or 50 tons/h.

The decisions supported with all measurement information are process and quality management, special actions, such as grade changes, recovery from process upsets and decisions about rejecting product batches, and configuration of the measurement information system itself. Furthermore the most costly laboratory measurements are typically used for three purposes: firstly to validate on-line (scanning) sensors, secondly to manage such quality parameters as paper strength that cannot be estimated on the basis of on-line measurements, and thirdly to decide whether a machine reel confirms to quality specifications or needs to be rejected. Each of these purposes sets requirements on how much uncertainty may be tolerated in the quality estimate for nth variable, σ_n^2. Lot of information is derived from the process and used in many ways, but it is poorly known, how and if the operators exploit all the information available. Therefore there has not been systematic work on optimizing the measurement activities. This may lead to situation where some measurements are carried out without purpose, only by habit, and the common practice continues to be to measure all quality parameters at regular time intervals. Obviously this is costly and rigid and may limit possibilities to measure those quality parameters that would be most important for overall quality management and thus for decision support. However, if end user information requirements and constraints on uncertainty of the measurements are made explicit, the optimal arrangement of the measurements can be determined (Grén, 2006).

In this paper we discuss systematic approaches to optimizing laboratory practices. In chapter 2 we define and solve the general case of scheduling laboratory measurements if the covariance of quality parameters and the costs of all possible measurement combinations are known. The solution minimizes the costs under the constraint that the uncertainties of quality parameters are below a prescribed value. In chapter 3 we discuss how the workload at laboratory can be minimized with on-line quality analyzer measurements. The method is demonstrated for mill case of managing optical properties of paper. Also this analysis is based on the covariance of measurements. Our practical results show that covariance develops slowly in time and therefore the scheduling must be accompanied with an updating scheme. Chapter 4 discusses our methods and their applications.

2. OPTIMAL MEASUREMENT SCHEDULE AT LABORATORY

In this chapter we define and solve the general case of scheduling laboratory measurements, what to measure and when.

Quality properties of paper are correlated. Therefore measuring one of the properties will provide information about the other as well. As time evolves the uncertainty about quality increases, if no measurements are made. Therefore the uncertainty in quality information has dynamics of continuous increase between measurements and stepwise decrease when measurements are made.

Quality is managed through decisions and actions about the production process. In general, the decisions need to be made at time instants independent on the measurement instants. The performance achievable depends on how uncertain the quality information is.

The cost of making measurements is not simply the sum of the costs of individual measurements. A group of quality parameters can be measured with a single device with small additional cost to that when measuring only one of the parameters in the group. But then again, there may be two quality parameters at the simultaneous measurement of which is higher than the sum of costs of individual measurements because of, for example they require the same sampling, equipment or personnel resources.

In this chapter we formalize these considerations in a somewhat idealized case, and solve the consequent optimization problem. For closely related analysis, see for example (Mehra, 1976 and Bicchi & Canepa, 1993).

We assume that on the basis of history data we know exactly the correlation between quality variables, and that the joint probability density is Gaussian distributed:

$$X \sim N(\mu, \Sigma^{(ap)}) \quad \text{(Eq. 1)}$$

Here X is the vector of quality properties, μ its expectation vector and $\Sigma^{(ap)}$ the covariance matrix describing the correlation between quality variables

We assume that the quality measurement history or a specific measurement campaign provides us the a priori probability density function for quality variables. Then our quality information dynamics consists of two mechanisms: continuous degradation towards the joint a priori probability density of quality, and occasional updates when new measurements are made.

We describe the degradation with probability density function dynamics, the Fokker-Planck equation, and the updating with probabilistic description of measurement and by applying Bayesian combination of earlier degraded information and of fresh measurement information. With this information dynamics we are able to assess the quality uncertainty at any time with any measurement schedule. So, it is assumed that between the measurements quality vector evolves according to a linear stochastic differential equation (Ornstein-Uhlenbeck process, OUP) and the equation is chosen so that the quality information approaches asymptotically the distribution of (Eq. 1) if no measurements are made.

Let us now assume that at time t, the uncertainty about quality is described with a normal distribution with mean $\mu(t)$ and covariance matrix $\Sigma(t)$. By solving OUP we find that at time $t+\Delta t$ without new measurements made, the uncertainty continues to be described with normal distribution with mean and covariance given as:

$$\mu(t+\Delta t) = \mu(t) + (\mu - \mu(t))(1 - \exp(-\Sigma^{(ap)^{-1}} D \Delta t / 2))$$
$$\Sigma(t+\Delta t) = \Sigma(t) + (\Sigma^{(ap)} - \Sigma(t))(1 - \exp(-\Sigma^{(ap)^{-1}} D \Delta t))$$

(Eq. 2)

where D is the symmetric diffusion matrix describing the rate at which the quality information deteriorates.

Assuming that measurement uncertainties are normally distributed and that a subset $X^{(2)}$ is measured while $X^{(1)}$ is estimated, we derive the recursion for covariance matrix of measurement information about $X=[X^{(1)} X^{(2)}]$ at $t+\Delta t$ and t to be

$$[\Sigma(t+\Delta t)] = \left[\left[\Sigma(t) + (\Sigma^{(ap)} - \Sigma(t)) \left(1 - \exp\left[-\Sigma^{(ap)^{-1}} D \Delta t\right]\right)\right]^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & C_{22}^{-1} \end{bmatrix} \right]^{-1}$$

(Eq. 3)

where Σ (and $\Sigma^{(ap)}$) are covariance matrices for the quality information (respectively a priori information) about X ,whereas C_{22} is the covariance matrix of measurement uncertainties in $X^{(2)}$. D is the symmetric diffusion matrix for the OU process.

Given the quality information dynamics, Eqs. (2-3), and let k be a binary vector indicating at each time instant i the measurements made, we may then formulate the optimal measurement schedule problem as:

$$\min_{\{\bar{k}(i)\}} \sum_{i=1}^{I} c(\bar{k}(i))$$
$$s.t.$$
$$\Sigma(t)_{nn} < \sigma_n^2 \quad \forall t, n \qquad \text{(Eq. 4)}$$

where c($\bar{k}$) is the cost of making measurement $\bar{k}$.

In the case of Gaussian distributions, the uncertainty is independent of measurements made, and thus solving Eq. (4) defines an optimal policy. This combinatorial problem can be solved for example with simulated annealing.

2.1 Simulated Annealing

Simulated annealing is a generic probabilistic meta-algorithm for the global optimization problem, namely locating a good approximation to the global optimum of a given function in a large search space.

The name and inspiration come from annealing in metallurgy, a technique involving heating and controlled cooling of a material to increase the size of its crystals and reduce their defects. The heat causes the atoms to become unstuck from their initial positions (a local minimum of the internal energy) and wander randomly through states of higher energy; the slow cooling gives them more chances of finding configurations with lower internal energy than the initial one.

By analogy with this physical process, each step of the simulated annealing algorithm replaces the current solution by a random nearby solution, chosen with a probability that depends on the difference between the corresponding function values and on a global parameter T (called the *temperature*), that is gradually decreased during the process. The dependency is such that the current solution changes almost randomly when T is large, but increasingly downhill as T goes to zero. The allowance for uphill moves saves the method from becoming stuck at local minima. (Aarts & Korst, 1989)

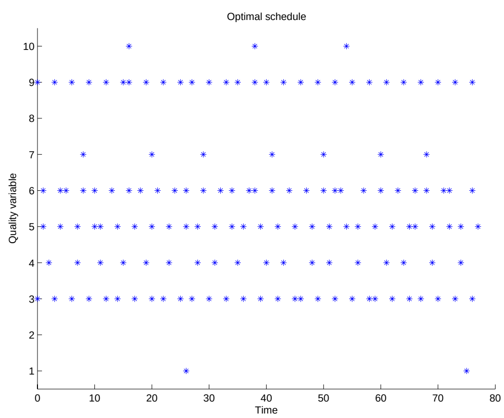

Figure 1. An example of optimal schedule obtained through solving Eq. (4). Star denotes a measurement made. Note that variable 2 and 8 is not measured at all as the accuracy requirements for them can be obtained from measuring other quality variables.

Figure 1 shows an example of the policy for 10 measurements over a horizon of 80 time instants, and thus the search space has 2^{800} possible schedules. Figure 2 shows the corresponding uncertainty in one of the variables.

3. WORK LOAD MINIMIZATION AT LABORATORY

In this chapter we describe one practical application of the principles described in previous chapter. At paper mills the active quality control is broke management and apportioning raw materials. The most important special action in quality management is detecting off-specifications products to be rejected or downgraded. The validation and calibration of on-line quality sensors with other quality measurements is the main configuration decision.

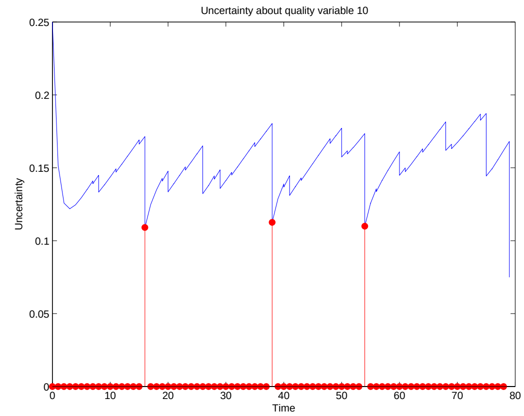

Figure 2. Uncertainty variation of quality variable 10. Measurement instants are indicated as dots.

The accuracy constraints for quality information, (Eq. 4), can be derived from analysis of these decisions: how data is actually used in each of the cases. There are only very few mills that have carried out such an analysis, and no mills that have applied decision analysis to specify how accurately the measured parameters must be known.

The following example of finding the measurement setup that optimizes the work load between an automated quality analyzer (27 quality measurements) and laboratory analysis (14 quality parameters) is based on analysis of quality measurement data from a paper mill, several months from both setups. The analysis concentrates on one paper grade only. The aim of this optimization problem is to minimize costly and tedious laboratory analysis measurements by using mutual information between automated quality analyzer and laboratory analysis.

As a particular case we consider the management of the key optical properties of paper: brightness, opacity and L-a-b colour coordinates. The optical quality specifications of printing paper grades set by customers are tight as the visual appearance of printed products hinges on these quality parameters. The optical properties are well standardized and there exist laboratory devices of high accuracy to measure these parameters. However, such laboratory activities are labour intensive and also require investing in devices and systems. Laboratory measurements can never be made with high enough frequency to give a sufficient understanding of optical quality variations within a customer batch. These properties can be measured with automated quality analyzers that have investment costs of similar magnitude but the operational costs per sample are much lower. The practices of combining laboratory analyses and automated quality analyzers has been developed over time into their present form, and it may be questioned whether they are close to optimal.

Based on correlations between measurements we can find optimal quality model to substitute laborconsuming laboratory measurements. Then we can reallocate the resources for making such kind of tedious laboratory measurements.

So, following linear optimal estimation we have two quality parameters groups $X = [X^{(1)}\ X^{(2)}]$ and measuring $X^{(2)}$ (automated quality analyzer) generate the estimates for $X^{(1)}$ (laboratory measurements). Every quality parameter in $X^{(1)}$ has the largest allowable uncertainty (constraint of Eq. 4) and every quality parameter in $X^{(2)}$ has individual costs (see Eq. 5.). Objective was to find correlating quality parameters between these two measurement methods – laboratory analysis and automated quality analyzer – and to find the group of quality parameters that can be estimated using measurement results from automated quality analyzer only, thus reducing the work at laboratory. After that labour intensive laboratory analysis can be focused to those quality parameters that cannot be estimated using this model and are important for overall uncertainty management and thus for decision support.

Quality parameter data was divided into two parts; the first part (three months) was used as identification data for establishing the covariance based quality model, and the second part (the following months) as validation data.

Identification data was used to generate a quality model for all of 14 laboratory analysis parameters and to find the optimal set of automated quality analyser measurements. This quality model was constructed using stepwise regression, which chooses between regression models with an automated sequence of F-tests. A forward selection applied here starts with no variables in the model, tries out the variables one by one, and includes those in order of statistically significant (Draper & Smith, 1998). Stepwise procedure was modified such that analysis was aborted when R^2-value reached 0.8.

Other modification was the measurement costs of the quality parameters $X^{(2)}$. The cost parameters were implemented to correlation and partial correlation calculations of stepwise regression. Idea was to penalize expensive measurements and lower their correlations to the target value. This was defined as:

$$C_K = \frac{1}{K} * C^2 \qquad \text{(Eq. 5)}$$

where K is cost value that is normalized between 0.5 and 1 so that value 1 is the most expensive measurement, C is correlation coefficient and C_K is the correlation coefficient with cost effect.

As time evolves it is self-evident that the uncertainties of estimates of quality model increase. We assume random walk (Bar-Shalom, Rong & Kirubarajan, 2001), with diffusion constant D and this is important tuning parameter of our system, telling how fast quality information deteriorates (Latva-Käyrä, 2002). We monitor when uncertainty of any part of the quality model exceeds the pre-given limit. These limits is partly based on careful examination of process data, partly information from the mills automation system and partly based on silent information from the operators. Yet it has to be considered if the limits are optimal or not.

After any limit is about to exceed the quality model parameters needs to be updated and corresponding measurements are made. Of course between these time instants the laboratory analyzer measurements can be made but here the effects are left out. The results tell us when and which laboratory measurements need to be done.

Figure 3 shows policy for 14 laboratory measurements after the quality model with D = 0,001 is initialized and figure 4 shows the corresponding uncertainty variation of quality variable 11, which is modeled using six quality analyzer measurements and figure 5 shows uncertainty variation of quality variable 14, which is modeled using only one quality analyzer measurement. Solid thin (blue) lines represent the estimate uncertainty with D (0,001), thick black lines the largest allowable uncertainty, black circles represent measured and (red) crosses estimated values and black dotted line represents the target value.

Figure 4 presents the uncertainty variation of quality variable 11 and it shows nicely how this quality model works. In the case of process disturbance of course the uncertainties of the estimates of the quality model exceeds the limits but then the disturbance has to be corrected and at the case of process change the quality model needs to be updated. An apparent asymmetry in the estimate uncertainty occurs from variation of the estimate.

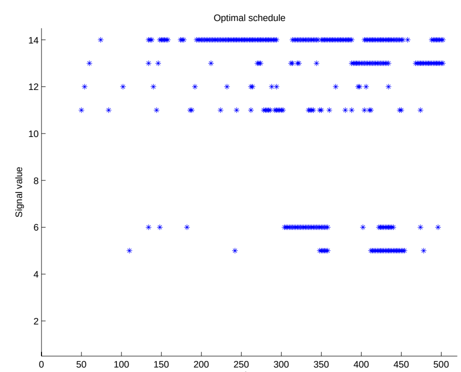

Figure 3. An example of schedule.

Star denotes a measurement made. Note that some variables are not measured at all after the initialization on quality model as the accuracy requirements for them are within broad limits.

Looking figures 3, 4 and 5 it shows that some quality variable can be modelled accurate enough but for example quality variable 14 cannot be modelled with this quality model. There is only one analyzer measurement correlating enough with this quality variable. It is then evident that more effort has to be put measuring this kind of variables, if this quality variable is important part of the overall uncertainty management.

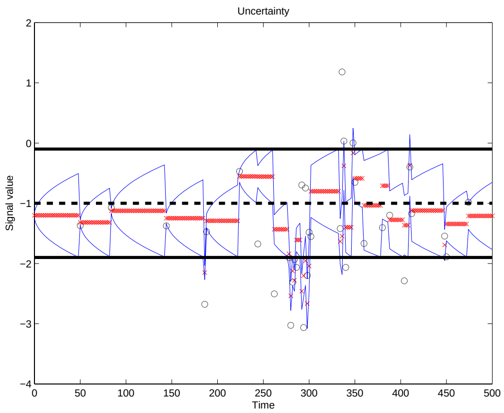

Figure 4. Uncertainty variation of quality variable 11. Six analyzer measurements are being used at the modeling.

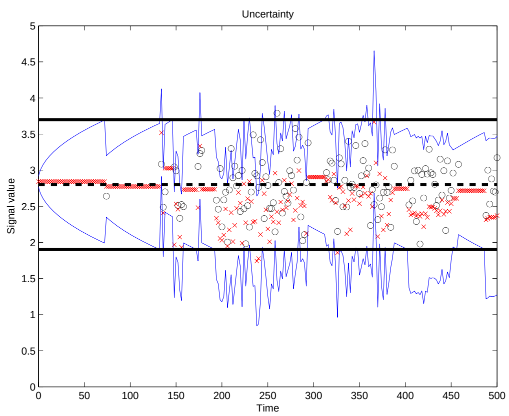

Figure 5. Uncertainty variation of quality variable 14. Only one analyzer measurements is being used at the modeling.

The degraded laboratory information and the information based on analyzer results can be fused to reduce the number of laboratory analyses. The analysis of the validation shows in general that the predictive power of laboratory analyzer, using the same parameters has degraded substantially. Therefore, occasional laboratory measurements are needed to dynamically validate the coefficients. The frequency of needed laboratory measurement is, according to our analysis, much lower than current practice (Gren, Konkarikoski & Ritala, 2007).

4. DISCUSSION

The optimal measurement system and balance between using the quality analyzer and laboratory analysis can be achieved with the methods represented above. This means that the uncertainty of state information is kept below the pre-specified level while minimizing the costs of measurements. Therefore the current practice, when all quality parameters are measured regularly, can be replaced with more efficient practice by measuring only the crucial parameters when needed. In practice the most costly laboratory measurements are replaced with estimates found via quality analyzer. Hence the measurement costs can be reduced and resources directed to measure more accurately the quality parameters that would be most important for overall uncertainty management and thus for decision support.

Methods represented above assume explicit made end user information requirements and constraints on uncertainty of the measurements. Finally information from measurements, soft sensors and simulators generates value through improved decisions (Raiffa & Schlaifer, 1960), because the uncertainty about the present and/or future state of the process has been reduced. However, the amount of value generated depends on the goal set by the decision maker, including the decision maker's attitude towards risk. Process operators and engineers are rather unfamiliar with the concept of uncertainty and hence uncertainty of consequences is dealt with rather implicitly when making decisions (Ritala et al, 2004). Therefore the set limits for uncertainties have to be considered carefully before implementing the represented methods in practice, although the process of finding these limits can be very informative and useful.

More effective would be to use fast, but often inaccurate, on-line measurements to estimate the laboratory measurements and to use laboratory analyzer measurements when validating the estimation models. The analysis with the presented framework – formal or expert derivation of accuracy of quality information required, optimal estimation analysis of opportunities to replace labour-intensive measurements by estimates, and dynamic degradation analysis to derive frequency of measurements – however, pinpoints critical measurements and concentrates more effort on them. In most cases the analysis process itself is of high importance: it provides a shared and documented view on performance requirements for the quality measurement activities; the accuracy of information, the availability and the costs related. Knowledge about the engineered accuracy and reliability of the measurements increases the operators trust at the quality parameters, thus supporting and improving decision making.

5. CONCLUSION

Measurements are uncertain and estimates of real world derived from them are uncertain. Therefore operational decisions are always made under uncertainty. The value of measurement information is determined by how much the decisions can be improved based on it.

In this paper we have shown a method how to find optimal measurement schedule, what to measure and how often. We have discussed our results with real mill quality data with emphasis on optical and strength properties of magazine papers. We have showed how the models required for optimal scheduling are established and maintained, and how a practical cost structure of operating the laboratory measurements is analyzed.

Establishing the limits for acceptable uncertainty about quality has in practice turned out to be the most difficult task when applying our approach. In this paper we have discussed how the acceptable uncertainty can be determined by analyzing the decisions support with laboratory quality data and how human factors affect this analysis. At some paper mills where the project of finding acceptable uncertainty limits has been tried out, the project itself has been informative and useful, because of brainwork about the process, but here is still lot of work to do.

Our future research tasks include implementation of covariance matrix of laboratory measurements so that every time when laboratory analysis is made maximum amount of information is exploited.

6. REFERENCES

Aarts E., & Korst J. (1989). *Simulated Annealing and Boltzmann Manchines.* Wiley, 0-471-92146-2, USA

Bar-Shalom Y., Rong Li X. & Kirubarajan T. (2001). *Estimation with Applications to Tracking and Navigation*. Wiley, 0-471-41655-X, USA.

Bicchi A. & Canepa C. (1994). Optimal design of multivariate sensors, *Meas. Sci. Technol.* 5 (1994), 319-332

Draper N.& Smith H.(1998) *Applied Regression Analysis*, Wiley Series in probability and statistics. 0-471-17082-8, USA.

Grén, J. (2006). *Systemizing the set and frequency of quality measurements in quality control.* Diploma Thesis, Tampere University of Technology, Tampere.

Gren J., Konkarikoski K. & Ritala R. (2007). Optimal Measurement Policy for Decision Making: A Case Study of Quality Management based on Laboratory Measurements. *Metrology and Measurement Systems*, Vol. XIV, 1/2007, 165 – 180.

Latva-Käyrä K. (2002). Data deterioration: using laboratory measurements for dynamic calibration. *Proceedings of Information, Decision and Control.* pp. 141-144. 0-7803-7270-0, IEEE IDC 2002, Adelaide, Australia.

Latva-Käyrä K. (2003). *Dynamic Validation of On-line Consistency Measurement.* Ph.D. thesis, Tampere University of Technology, 952-15-1101-X, Tampere.

Mehra, R.(1976). Optimization of Measurement Schedules and Sensor Designs for Linear Dynamic Systems, *IEEE Trans. On Automatic Control*, 21, (55-64).

Raiffa H. & Schlaifer R.(2000, originally published 1960) *Applied Statistical Decision Theory*. Wiley Classics Edition, 0-471-38349-X, USA.

Ritala R., Belle J., Holmström K., Ihalainen H., Ruiz J., Suojärvi M. & Tienari M. (2004). Operations Decision Support based on Dynamic Simulation and Optimization. *Proceedings of Efficiency*, pp. 55-62, PulPaper 2004, Conferences, - Energy, Coating, Efficiency, Helsinki, Finland.