Value assessment of measurements in large measurement information systems

Heikki Jokinen, Risto Ritala

Tampere University of Technology, Institute of Measurement and Information Technology,
P.O.Box 692, FIN-33101 Tampere, Finland, fax: +358 3 3115 2171, email: firstname.lastname@tut.fi

Abstract **-** Modern industrial production processes are rich sources of measurement data, acquired in many forms and time periods. The integration of different forms of data and information has become more and more important in process industry. Decisions needed in managing processes in an optimal way force us to utilize the information in a way that maximizes its value. This paper discusses the value assessment of measurements and measurement information in large industrial systems. In particular, probabilistic description of information and the value of uncertainty and information in measurements concerning the performance of operational decision making are discussed.

I. Introduction

What is the value added by introducing a new sensor to an industrial process? What is the value of data analysis method supporting operators' decision making? How to pinpoint sensors and laboratory measurements that add value less then their maintenance/operation costs, so that these should be removed? What is the value that a new controller or more advanced control scheme adds to the operation? These are the questions that are faced when designing new processes or process rebuilds. However, the decisions made about selecting measurements are currently based on folklore or perception rather than systematic cost/benefit analysis. This is because the theoretical approach to these questions is largely missing, and because of the complexity of the information systems.

Industrial processes – such as paper and pulp mills or power plants – contain a few thousand measurement data sources sampled at intervals lasting from fractions of a second to a few minutes. The data acquired and stored in history data bases must be combined in order to obtain the estimates of system parameters of interest, optimal in terms of their reliability. The combination of data is necessary as both the systems and measurement processes are stochastic. Measurement data as such does not support decision making, but it has to be transformed into information [1] through the methods of data analysis by integrating data and prior information from different sources, in order to be able to estimate optimally the piece of information required in decision making. The new data is also used in updating prior information. Finally the decision making process turns this information into actions. Decision making is an optimisation task that aims at maximizing the system performance measured according to an explicit or implicit criterion, under stochasticity due to the uncertainty about the states of the process. Fig. 1 summarizes the process from measurement to actions.

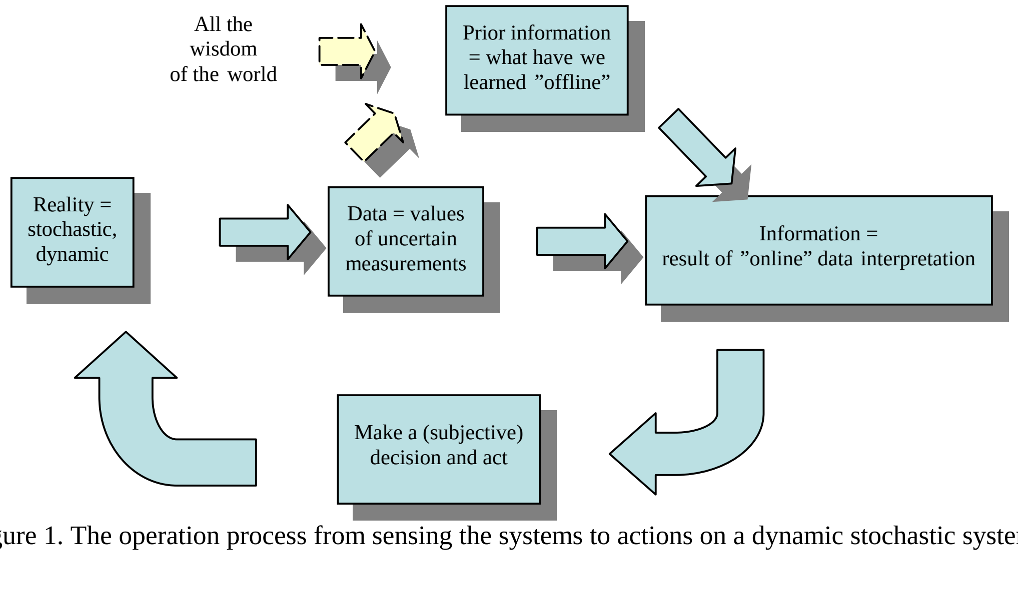

Figure 1. The operation process from sensing the systems to actions on a dynamic stochastic system

II. Value assessment of measurements and information

The design of material and energy flows in industrial processes is based on a careful cost-benefit analysis and optimisation of process performance [2]. The design of information processing systems, however, is rather heuristic due to the difficulty to assess the value of information or the cost of uncertainty. There is no systematic way to estimate, for example, the value added by a new sensor introduced to the process, or the value of data analysis supporting the operators' decision making. Neither is the removal of sensors and laboratory measurements adding value less than their operational costs easy to justify. This is because there is a lack of theoretical approach being capable in fulfilling these needs and performing a systematic analysis for the costs and benefits of measurements. However, such design and redesign decision are made daily in process industries. Therefore, we address the question: how much better decisions can be made when a new piece of information is added to the system, or when the uncertainty of information is reduced.

The key questions in design of a measurement information system supporting decision making are:

- What is the value of a measurement or a piece of *a priori* information measured in terms the achievable performance of the system through decision making?
- How much of performance is lost due to that both the measurement data and *a priori* information is uncertain?

Our aim for the research is to be able to combine several theoretical approaches into a framework that can be used to assess the value of measurements and information with applications to relevant industrial design problems. In spite of developing the approaches themselves, the challenge of the research is in introducing the strategy of co-operation between the approaches.

The theoretical approaches applied to the assessment of value of information are:

- *probabilistic*, (Bayesian), *and dynamic description* of information obtained through measurements about the target process (reality)
- *information theory* (entropy) for describing the content of information and changes in it as new data and information becomes available
- decisions under uncertainty understood as *dynamic stochastic optimisation* tasks
- objectives of decisions analysed in terms of utility functions and other concepts of *statistical decision theory*.

A. Probabilistic description of information

Industrial production processes are dynamic, and thus any decisions and actions will affect the future over several time intervals of decisions. Our knowledge about the state of the process is incomplete because of incomplete measurement of the process state and because of measurement uncertainty. Thus we have to describe the information about the state of the process – the reality – in probabilistic terms [3]. Any additional information will change this information according to Bayes' theorem.

Bayes' theorem [4] is known as a method used in updating probabilities based on acquired information. This can be illustrated with the simplest form of the theorem

$$P(B \mid A) = \frac{P(B)P(A \mid B)}{P(A)}, \qquad (1)$$

where *A* and *B* denote the two possible events and $P(A \mid B)$ is the likelihood given to *B* by *A*. *B* is the event of interest, and $P(B)$ denotes the *prior* probability of its occurrence. After observing the occurrence of *A*, the *posterior* probability of *B* with *A* known to have occurred is $P(B \mid A)$. This probability is thus updated from the prior probability by multiplying it with the ratio $P(A \mid B)/P(A)$ explaining the probability changes after acquiring new information.

In terms of random variables, we can rewrite the Bayes' theorem as

$$f(\theta \mid x) = \frac{f(\theta)\, f(x \mid \theta)}{f(x)}, \qquad (2)$$

where x and θ denote quantitative data and the parameter of interest, respectively, and $f(\cdot)$ is the density function. The new information acquired with the data *x* is used in updating the distribution of θ and producing the posterior distribution $f(\theta \mid x)$.

In probabilistic measurement information theory we are interested in the state of reality [5], now denoted as θ given that when we have measured the state of reality we have obtained, *x*. The probability density of measurement *x*, given the state of reality is θ, is an intrinsic property of the measurement system or device, $f(x \mid \theta)$. Hence (2) is interpreted to yield the quantity of interest, the probability density of reality with given a priori about reality $f(\theta)$, given properties of measurement system and the measurement reading, *x*. For example, if the measurement system is unbiased, has uncertainty σ, can be described as normally distributed, and there is no prior information about the state of reality, our information about the state of the reality after measurement is

$$f(\theta \mid x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-(x-\theta)^2 / 2\sigma^2\right]. \tag{3}$$

Eq. (2) is the way to handle the information about the state of reality as additional information becomes available, and as information obtained with earlier measurements degrades over time.

The Bayesian inference is then performed based on the posterior distribution of θ with the help of useful presentations like distribution functions. These are especially needed to reveal the possible multimodal forms of the posterior distribution. Furthermore, in practice, the cases are involved with multiple variables leading to bivariate or multivariate distributions. The decision is usually made based on the modes of the distribution, i.e. the points revealing higher posterior probability.

B. Information theory

Mathematical information theory provides well-known measures of content of information. Information content is expressed with *entropy* [1], defined as

$$H(X) = -\sum p(x) \log p(x), \tag{4}$$

where $p(x)$ denotes the probability of the variable *x*. Another important quantity is *mutual information* that measures the dependence of two variables, i.e.

$$I(X;Y) = \sum p(x,y) \log \frac{p(x,y)}{p(x)\,p(y)}. \tag{5}$$

Mutual information can be understood to measure the amount of information transferred from target *x* to measurement *y*.

As such, mathematical information theory is basically a theory of signal transmission. Its main care is to ensure a set of signals produced at one point to be reproduced at another point with high accuracy. According to mathematical information theory, whichever set of signals consists of information, despite of its meaning. That is why, in order to be able to assess also the meaning and the value of information, the processing of semantic and especially pragmatic aspects of information are needed. Acquired information does not contain any value to be utilized without being to be able to reduce the uncertainty concerning the states the decisions are based on, and to increase the amount of knowledge that allow us to save resources and increase our lot.

The value of information is generated in decision making and is not directly linked to information content. Still, the information is needed in supporting decisions. We need a systematic way to analyse the additional performance gained due to information in designing optimal flows of information in process industry [2]. The main components of information needed are (1) the current state of the process, (2) the prediction of the process state given a scenario of future actions, (3) prediction of external factors affecting the system, and (4) predicting external factors affecting the measure of performance.

C. Dynamic stochastic optimisation

Operational decision tasks in process industries can be understood as dynamic optimisation tasks, as has been discussed recently concerning pulp and paper industry [6]. The uncertainty about the present state and about the external factors affecting the process adds stochasticity to the problem. Let us first neglect the stochastic aspects. Then the decision optimisation consists of the following [7]: optimisation time horizon, instantaneous objective (to be summed over the optimisation horizon), end point objective, set of possible actions, inequality constraints, and model to predict to future evolvement of the system.

Optimisation time horizon has a strong effect on the optimal solution. Process operators, engineers, mill managers, and corporate executive officers make decisions that are related to one another but with different horizon.

Primary instantaneous objective is the efficiency or profitability of the production line. Often it is rather difficult to relate the actions considered directly to profitability and thus intermediate objectives need to be defined. The role of end point objective is to ensure that at the end optimisation horizon, the operating is not a poor starting point, e.g. all storages empty.

Set of possible actions defines all the decisions that can be made, irrespective of the outcome. Inequality constraints specify the operational regions that must be avoided because of equipment constraints, safety issues or operational policies. When implementing decision support, the operational policies should be carefully analysed: which to keep, and which to relax? Some of the operational policies are in place only to ease the decision making in this complex environment. If the decision support system to be implemented supports the same decision making, the operational policy may be relaxed and thus room for improved profitability or efficiency is opened.

Due to uncertainty, the instantaneous and end point objectives are stochastic with probability distributions derivable from distributions describing the uncertain factors. The decision maker has to define his/her attitude towards risk versus expectation value of performance. Furthermore, the (hard) inequality constraints will become probabilistic, too.

As an example of stochasticity [6], the time needed for a grade change at a paper machine is illustrated with two probability distributions introducing two possible approaches to perform a grade change, see Fig. 2. The first approach has lower average grade change time but higher probability of losing too much time for the change, while the another approach will produce higher average with smaller uncertainty.

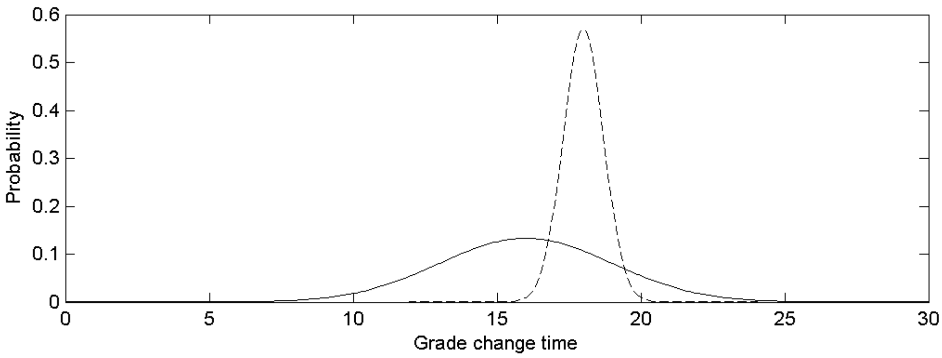

Figure 2. Probability distributions of two grade change approaches.

D. Decision making

The statistical decision theory is concerned with the problem of making decisions in the presence of statistical knowledge about uncertainties involved [8]. Statistical prior knowledge may be deduced by making experiments. However, this knowledge is not sufficient in making decisions, but also knowledge about the consequences of the decisions, and the prior information coming from the past experience, applying subjective knowledge and skills, is needed. The knowledge concerning the consequences may be quantified by determining the value of each decision. Understanding decision as optimisation tasks is a natural approach that has been utilized in e.g. model-predictive automatic control. However, when deciding under uncertainty, the cost of risk is to be included. Decision making has subjective features to it: the cost of risk is most commonly subjective. Utility theory provides a way to analyse the cost of risk and the added value of information [4], [9]. Utility is a function of all possible circumstances and constraints affecting the decision, especially of new information and actions taken. The utility can be assessed to every combination of circumstances at any time, and the decision maker applies it in trying to make an optimal decision.

The value of information provided by the data measured may be assessed as the difference expected between the utilities corresponding to optimal decisions after and before the information gained. As presented in [4], the value of information is connected to delaying the decisions until some new information is available. It is expected that the piece of information changes the utility and thus the decisions made after the new information produce more value compared to decisions made at once.

However, in our case of dynamic systems the overall assessment of values must also take into account the cost of delaying the decisions that may diminish the benefit produced by the new information through degradation of present information and through the evolvement of the system during making the additional measurement.

In the context presented the value of information is connected to making a new measurement or a test to gather new information. Rather than that, our aim is to assess the value of information gained in the long-term with investing on totally new measurement unit/data analysis tool or discarding an existing one.

E. Example: the value of measurement in choosing a grade change approach

Many industrial processes are designed for multigrade production. The minimization of grade change time is an important economic factor. For example, paper mills produce different grades of paper with specific mixtures of chemicals. A grade change means that the chemical state of the process has to be adjusted to meet the demands of the new paper grade, i.e. the disturbing chemicals have to be removed. In this example, there are two approaches to perform a grade change: an 'active' one with dosing chemicals that suppress the disturbing chemicals, and a 'passive' one that is performed without additive chemicals. The grade change time *T* depends on the concentration of the disturbing chemicals; however, by using the 'active' approach the time needed may be known with small uncertainty, while the distribution of possible grade change times is wide when the 'passive' approach is applied. This is illustrated in Fig. 2, the distribution of grade change times for the 'passive' approach (Approach 1) is denoted with a solid line and the 'active' approach (Approach 2) with a dashed line. The distributions are expected to be normal. However, if we measure the concentration *x* of the disturbing chemicals in Approach 1, we may diminish the uncertainty and produce a more accurate estimate for the time *T*. Then depending on the value *x* measured, we may make a more educated decision whether to choose the active or passive approach. When we choose the 'passive' approach based on measurement *x*, we denote it Approach 1^*.

Without measuring the concentration, the change times for Approach 1 may be estimated as $N(\mu_1,\sigma_1)$-distributed, while having made a measurement *x*, the distribution is $N\left(\mu_{1^*},\sqrt{\sigma_{1^*}^2+\sigma_{\text{meas}}^2}\right)$, with $\sigma_{1^*} << \sigma_1$, $\mu_{1^*}(x) = Ax+B$ by expecting a linear dependency between concentration and grade change time, and σ_{meas}^2 denoting the scaled measurement error. Additionally, if the decision about the approach has to be made later than immediately after having made the measurement, for example, due to computation of analysis results, the information is degraded due to delay Δt. The increase in uncertainty results in the standard deviation of the form $N\left(\mu_{1^*},\sqrt{\sigma_{1^*}^2+\sigma_{\text{meas}}^2+D\Delta t}\right)$, where *D* is a parameter describing the diffusion of uncertainty as an integrated white noise process. The distribution for Approach 2 is $N(\mu_2,\sigma_2)$ so that $\mu_1 < \mu_2$ and $\sigma_2 < \sigma_1$. However, $\mu_{1^*}(x)$ is either smaller or larger than μ_2, depending on the concentration *x*.

In order to be able to make the decision of choosing a proper approach, we compute the utility as a function of *T* as

$$u(T) = \mu + a\sigma, \qquad (6)$$

where *a* denotes the parameter for the cost of risk due to uncertainty. The utilities for the three approaches are as follows, 1: $u_1(T) = \mu_1 + a\sigma_1$, 1^*: $u_{1^*}(T) = Ax+B+a\sqrt{\sigma_{1^*}^2+\sigma_{\text{meas}}^2+D\Delta t}$, and 2: $u_2(T) = \mu_2 + a\sigma_2$. Without a measurement, the only parameter affecting the decision is *a*, resulting in choosing Approach 1, if $\mu_1 + a\sigma_1 < \mu_2 + a\sigma_2$. With measurement *x*, the decision also depends on x, σ_{meas}, and $D\Delta t$. Approach 1^* is chosen, if $Ax+B+a\sqrt{\sigma_{1^*}^2+\sigma_{\text{meas}}^2+D\Delta t} < \mu_2 + a\sigma_2$. The graphs illustrating the regions for optimal utilities of approaches 1 and 2, and 1^* and 2, as a function of *x* and *a*, are presented in Fig. 3 a) and b), respectively. The three lines in Fig. 3 b) illustrate the effect of different values of uncertainty $\sqrt{\sigma_{1^*}^2+\sigma_{\text{meas}}^2+D\Delta t}$, the uppermost line denoting the smallest value.

Computed with one value of uncertainty, the difference of utilities u_2 and u_{1*} at two fixed values of *a*, $a_1 = 0.5$ (–) and $a_2 = 1.5$ (– –), are presented in Fig. 3 c).

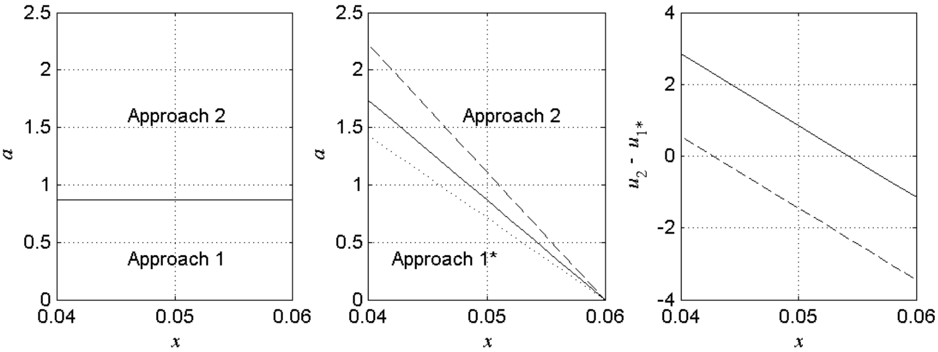

Figure 3. Decision regions for grade change approaches, a) without measurement of *x* and b) with measurement of *x* at three different values of uncertainties $\sqrt{\sigma_{1^*}^2 + \sigma_{\text{meas}}^2 + D\Delta t}$, and c) the difference of utilities u_2 and u_{1*} at two fixed values of *a*, 0.5 (–) and 1.5 (– –).

The measurement information has a clear value in aiding the decision making as a function of *x*. When both the concentration of disturbing chemicals and the cost of risk is high, we select Approach 2 independently of the value of the measurement, but, when we know, for example, that the concentration of disturbing chemicals is low (say, $x = 0.04$ and $a = 1.5$, see Fig 3 b), we are able to save raw material resources and achieve a smaller grade change time by selecting Approach 1^*. Furthermore, at low cost of risk levels (e.g. $a = 0.5$) and when the concentration is also low, we select Approach 1 (or 1^*), but at high values of concentration, Approach 2 might be selected.

In order to be able to answer the question of investing on a new measurement unit that produces *x*, we have to know the effect of *x* on the decisions related to the selected approaches and further, on change time in the long term. Due to this, we need to estimate the distribution of *x* for the whole life cycle of the unit, and estimate the expected value of change time over this distribution.

III. Conclusions

The aim of the research is to develop a framework for analysing the added value of information from measurements or prior information in the decision making process. The increase in performance in the whole system introduced by a new data source, and the amount of increased performance due to decreased uncertainty in measurements or prior information are the key concepts. The design of flows of information in industrial processes should be improved considerably with the framework developed.

References

[1] Cover, T.M., Thomas, J.A., *Elements of Information Theory*, John Wiley & Sons, Ltd, 1991.

[2] Biegler, L.T., Grossman, I.E., Westerberg, A.W., *Systematic methods of Chemical Process Design*, Prentice-Hall, London, 1997.

[3] Rossi, G.B., Crenna, F., "Probabilistic Uncertainty Evaluation Aiding Critical Measurement-Based Decision", *Proceedings of the IMEKO TC7 Symposium 'Measurement Science of the Information Era'*, p. 237-242, Cracow June 25-27, 2002.

[4] O'Hagan, A., Forster, J., "Bayesian Inference", in *Kendall's Advanced Theory of Statistics*, Vol. 2B, 2nd ed., Arnold, London, 2004.

[5] Rossi, G.B., "A Probabilistic Model for Measurement Processes", *Measurement*, Vol. 34, p. 85-99, 2003.

[6] Ritala, R., Belle, J., Holmström, K., Ihalainen, H., Ruiz, J., Suojärvi, M., Tienari, M., "Operations Decision Support based on Dynamic Simulation and Optimization", *Efficiency / PulPaper 2004 Conference*, Helsinki June 2-3, 2004.

[7] Bellman, R., *Introduction to Mathematical Theory of Control of Processes Vol 1: Linear Systems and Quadratic Criteria*, Academic Press, 1967.

[8] Berger, J.O., *Statistical Decision Theory and Bayesian Analysis*, Springer, New York, 1985.

[9] Bernardo, J.M., Smith, A.F.M., *Bayesian Theory*, John Wiley & Sons, Ltd, Chichester, 2000.