Novel mathematical and statistical approaches to uncertainty evaluation

A guide to decision-making and conformity assessment

Deliverable 3.3.1

Work Package WP3: Decision-making and conformity assessment

Conformance assessment of electrical energy meters investigated by risk analysis – a case study

Contents

Contents 2

List of Tables 2

List of Figures 2

Conformance assessment of electrical energy meters investigated by risk analysis – a case study ... 3

The specific risk analysis approach 4

Cost of rejection and cost of acceptance 5

Decision risks 7

Risk curves 9

List of Tables

Table 1 Statistical sampling plans in the Norwegian regulation. Indexes 1 and 2 refer to first and second part of the sampling plan ……..…….....3

Table 2 Example batch, metadata and results of testing. …….....4

List of Figures

Figure 1: Measuring errors for 76 samples in batch 2783 at three values of test current, compared to MPE. ……...5

Figure 2: Results of a Monte Carlo simulation of cost errors due to measurement errors in electric energy meters. The average cost error calculated from Eqn (3) is NOK 133,-.……7

Figure 3: P_C = 0% and P_{NC} = 100% at d = 22 and n = 76 for batch ID 2783. Maximum point for the distribution is at p_{max} = 22/76 = 28.9%.……..8

Figure 4: Probability density distributions for the true error rate, p, at different observations of the error rate, $\hat{p}$ and sample sizes, n …….……...10

Figure 5 Beta-distribution for the error rate using attribute sampling, calculation of probabilitie for conformance and non-conformance at first and second part of the sampling plan……10

Figure 6 For batch id 2783, PC and PNC for different values of d..……11

Figure 7 Average risk per instrument for producer (decreasing curves) and consumer (increasing curves) for the first (dashed lines) and second (unbroken lines) part of the sampling plan using watershed specifications. Explanations n_1 = 32, n_2 = 64, etc, refer to the sampling plans in the Norwegian regulation…….....11

Conformance assessment of electrical energy meters investigated by risk analysis – a case study

Accurate measurements of electrical energy consumption is important to ensure that the energy bills reflect the actual consumption. In addition to achieving a fair distribution of energy costs among the consumers, an accurate measurement of the consumption assists in levying a correct energy tax for the authorities, and it may motivate consumers to reduce their consumption in the most cost effective ways. Regulation quite reasonably places the responsibility for quality assessment of el-meters on the grid owners (producer side) rather than the consumers, with requirements directed both towards the performance of individual devices in the form of a maximum permissible error (*MPE)*, and towards entire batches in terms of the proportion of non-conforming items in the batch. Assuming a homogeneous batch, testing is carried out by acceptance sampling, which is a cost effective test regime compared to 100% control of all devices. The regulation prescribes retesting of batches at predetermined intervals. Predetermined test intervals in regulation range from 3 to 10 years.

The MID directive [EU Commission 2004 and 2006][1] is the basis for regulation of electrical energy meters in Norway. EN 50470-1/-2/-3[2] and OIML R 46-1/-2[3] describe the metrological requirements and tests. Additional requirements for statistical verification, using single or double sampling plans, are described in Welmec 8.10[4] and ISO 2859-1/-2[5].

Table 1 Statistical sampling plans in the Norwegian regulation. Indexes 1 and 2 refer to first and second part of the sampling plan.

Batch size	n_1	n_2	$d_{Accept1}$	$d_{Reject1}$	$d_{Accept2}$	$d_{Reject2}$
65 - 1 200	32	64	0	2	1	2
1 201 - 3 200	50	100	1	4	4	5
3 201 - 10 000	80	160	2	5	6	7
10 001 - 35 000	125	250	5	9	12	13

[1] EU Commission 2004 MID "Measurement Instrument Directive" *PE-CONS 3626/04* MID 2004/22/EC (2004), http://europa.eu/eur-lex/pri/en/oj/dat/2004/l_135/l_13520040430en00010080.pdf

[2] EN 50470:2006 "Electricity metering equipment (a.c.)" Part 1: "General requirements, tests and test conditions – Metering equipment (class indexes A, B and C)"; Part 2: "Particular requirements – Electromechanical meters for active energy (class indexes A and B)"; Part 3: "Particular requirements – Static meters for active energy (class indexes A, B and C)"

[3] OIML R 46-1/-2:2012 "Active electrical energy meters"; Part 1: "Metrological and technical requirements"; Part 2: "Metrological controls and performance tests"

[4] Welmec 8.10 "Measuring Instruments Directive (2004/22/EC): Guide for generating sampling plans for statistical verification according to Annex F and F1 of MID 2004/22/EC"

[5] ISO 2859:1985 "Sampling procedures for inspection by attributes" Part 1: "Sampling schemes indexed by acceptance quality limit (AQL) for lot-by-lot inspection" Part 2: "Sampling plans indexed by limiting quality (LQ) for isolated lot inspection"

Table 1 summarises the statistical sampling plan in the current Norwegian regulations of electricity meters, which are defined according to ISO 2859-2 “Sampling plans indexed by limiting quality (LQ) for isolated lot inspection”. The selection of accept and reject limits are based on certain quality levels, the AQL (Acceptance quality level) and the LQ (limiting quality). AQL is the quality level that is the worst tolerable, which can be accepted. MID requires that at 95% probability of acceptance (5% probability of rejection), the average fraction of non-conforming units shall be less than 1%. AQL is often referred to as the producers point of risk, because it shows that a sample from a conforming batch might by chance be falsely rejected. The LQ is the level of quality routinely rejected by the sampling plan. MID requires that at 95% probability of rejection (5% probability of acceptance) the average fraction of non-conforming units is less than 7%. LQ is often referred to as the consumers point of risk, because it shows that a sample from a non-conforming batch might by chance be falsely accepted.

It should be noted that the sampling plans in table 1 are selected from ISO 2859-2 and do not meet the MID-requirements for neither AQL- nor LQ-values. Examples on “best practice” for sampling plans in Welmec 8.10 refer to ISO 2859-1, and conform to MID-requirement for AQL-value, and, by a large margin, also the MID-requirement for LQ-value.

While the use of similar statistical sampling techniques are ubiquitous in quality assessment applications, the regulation does not explicitly consider any economic consequences of the decisions. Some critics has argued that the cost of replacing entire batches of el-meters is much higher than the cost of the measuring errors, which may well be true in certain cases. An alternative approach to decide on acceptance or rejection of a batch of el-meters involves an explicit cost calculation for each batch and attempts to strike a balance between the producers risk of rejecting and replacing a conforming batch with the consumers risk of accepting and keeping the measuring errors of a non-conforming batch.

The specific risk analysis approach

The analysis comprises essentially two steps: a computation of the probability of the actual failure rate, and a computation of the cost of the measurement errors. Both computations are performed from the observations.

To illustrate the approach the guide presents the results of testing an example batch, ID 2783, see table 2.

Table 2. Example batch, metadata and results of testing.

Metadata		Results of testing	
Batch ID	2783	Actual sample size, n_1^*	76
Type	Mechanical type	Number of failing units	24
I max	60 A	Failing units with respect to *MPE* (phi = 0 only)	22
Produced	1980 to 1982	Decision	Reject the batch
Number of units	6 730	Cost of replacement, pr unit	NOK 2 000,-
Sample size, n_1	80	Average annual cost of measuring errors, pr unit	NOK 133,-

The guide explains the calculations in a risk analysis approach to decision making. We calculate the costs of two opposite and mutually exclusive decisions, acceptance and rejection of the batch, and the probabilities of conformance and non-conformance, given the test result for the proportion of units in the sample failing to meet the requirements.

For a batch size of 6 730 units, a double sampling plan is prescribed, see row 3 in table 1. In the database, only 76 test results are available. Since more than 5 units fail to conform to requirement, the decision was clear before all sampled units were tested, and the test was curtailed.

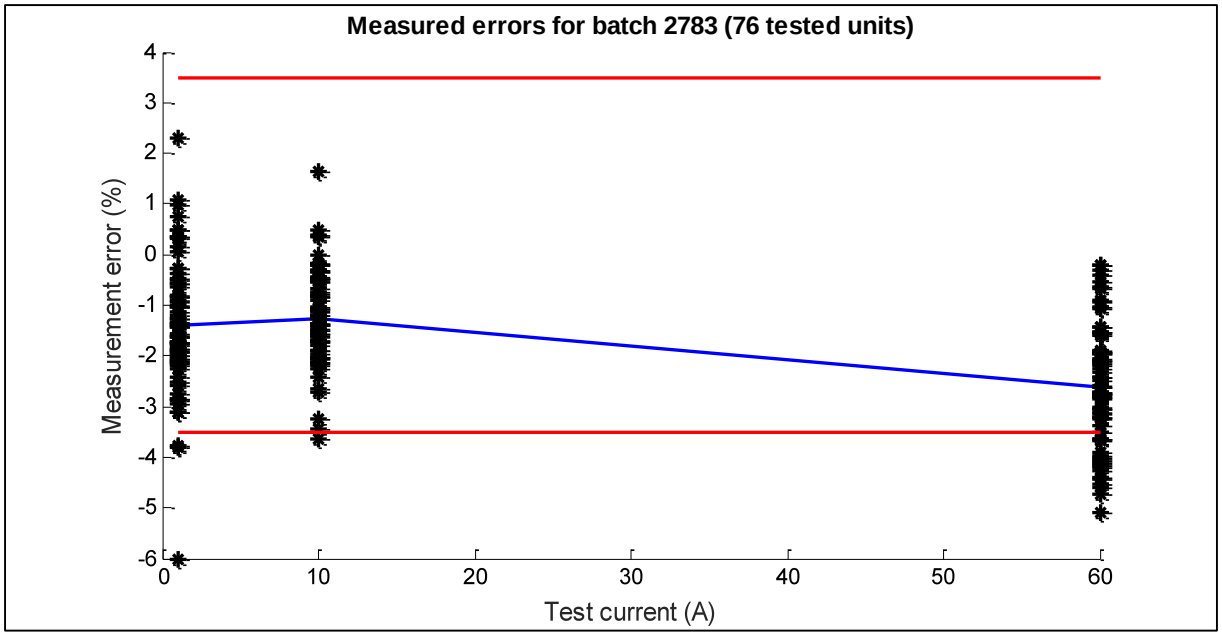

Figure 1. Measuring errors for 76 samples in batch 2783 at three values of test current, compared to *MPE*.

Cost of rejection and cost of acceptance

The cost of rejection is the cost of replacement of each unit in the batch, NOK 2 000,- , which comprises the installation and instrument cost.

The cost of accepting the batch is the sum of all costs of a continuation of erroneous measurements. The annual cost for a specific instrument *i* in the batch is calculated from equation (1):

$$c_i = c_i(W_i, P_i, e_i) = \sum_{q=1}^{8760} P_{iq}\left[W_{iq}\left(1 + \frac{e_{iq}}{100}\right)\right] \tag{1}$$

W_q = the actual electric power during hour number q

e_q = measurement error in % during hour number q

P_q = Price of electrical energy during hour number q

The annual cost error for instrument number i is given by Δc_i:

$$\Delta c_i = c_i(W_i, P_i, e_i) - c_i(W_i, P_i, 0) \tag{2}$$

For the hour-by-hour consumption a small sample of typical, actual user profiles were obtained from a large Norwegian utility company, (the W_q in Eqn (1)). From the same company, we also received information about the annual electrical consumption from more than 500 000 households in the eastern part of Norway. Spot price data (the P_q in Eqn (1)) on electric energy in Oslo can be downloaded from http://www.nordpoolspot.com.

With these data sets it is possible to apply Eqn (2) repeatedly in a Monte Carlo simulation:

(1) A utility meter is randomly drawn from the available test data.
(2) A random annual sum of electrical energy is drawn from the distribution of the 500 000 households.
(3) A random user profile is drawn and scaled in such a way that the annual consumption equals the value drawn in stage (2)
(4) The outputs from (1)-(3) are fed to Eqn (2) along with the price profile to produce a cost error.
(5) Everything is repeated a large number of times to expose the different electricity meters to different consumptions, and hence build a distribution of expected cost errors for the batch.

This calculation treats negative and positive measurement errors within the same utility meter fairly, the costs cancel out depending on the size of measurement error, frequency of use and price. A positive signed error is to the benefit of the producer, a negative signed error is to the benefit of the consumer. Both signs are equally important for *MPE* in the regulation, so this is also the case for cost errors. Independently of sign, the average annual cost error due to measuring errors for a randomly selected meter in the batch is calculated by equation (3):

$$Average\ cost\ error = \frac{1}{n}\sum_{i=1}^{n} |\Delta c_i| \tag{3}$$

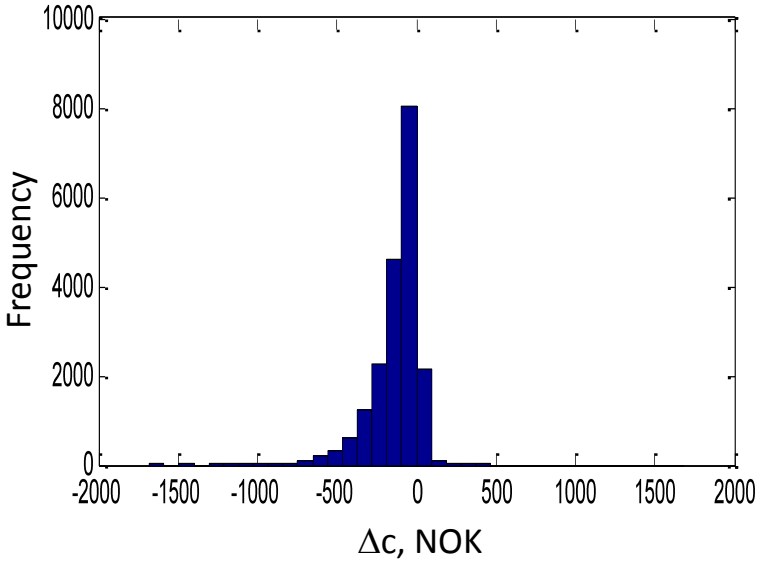

Figure 2: Results of a Monte Carlo simulation of cost errors due to measurement errors in electric energy meters in batch ID 2783. The average cost error calculated from Eqn (3) is NOK 133,-.

Figure 2 displays the result of a Monte Carlo computation of annual cost errors for the example batch id 2783. The distribution has a maximum close to zero, slightly on the negative side. It is also asymmetric with a heavy negative tail. The result of Eqn (3) is an average cost error NOK 133,-. If the decision is to accept the batch, the meters will continue measurements with such cost errors and a redistribution of the costs among consumers - with no other explanation than poor measurements - will continue. Longer test intervals will increase these cost errors linearly with time.

Decision risks

At this point, we know the costs connected to both decisions. We are now ready to proceed with calculations on the probability of conformance and the probability of non-conformance. The estimate of the error rate, $\hat{p} = d/n$, has an uncertainty, which needs to be evaluated.

The probability density function for the true error rate, p, is the normalized binomial distribution:

$$g(p;\hat{p}) = const \cdot p^d(1-p)^{n-d} = \frac{p^d(1-p)^{n-d}}{\int_0^1 u^d(1-u)^{n-d}\,du} \qquad (4)$$

p = true error rate
$\hat{p} = d/n$
d = number of non-conforming units in the sample
n = sample size

This expression is the beta - distribution with form factors α and β given by $\alpha = d + 1$ and $\beta = n - d + 1$.

In our example, 76 units were sampled and the number of units failing the requirement for *MPE* is d = 22. Figure 3 envisage the pdf for p given by the beta-distribution $g(p; \hat{p})$ in equation (4). Because of the large value for d, this is a symmetric distribution, quite similar to a normal distribution. Both the central value and the width of the distribution are determined by the form factors α and β, or d and n. $\hat{p} = d/n$ = 28.9% and σ = 5.1%.

The regulation defines criteria on d for acceptance and rejection in different sampling plans. These criteria translates into criteria on p for the conformance and non-conformance. $d \leq d_{Accept}$ translates into $\hat{p} \leq p_{Accept}$ and $d \geq d_{Reject}$ translates into $\hat{p} \geq p_{Reject}$. We could also have used a different approach, for example a single value, p^*, for all sampling plans: Accept if $\hat{p} \leq p^*$ and reject if $\hat{p} > p^*$.

d can only attain integer values, and to adjust for the limited resolution for the measurement of p, $\hat{p} = d/n$, we use watershed[6] specifications for the conformance and non-conformance regions. The watershed specifications for p_{Accept} and p_{Reject} becomes, for the current example:

$$p_{accept} = \frac{1}{n}\left(d_{accept} + \frac{1}{2}\right) \approx 0.033$$
$$p_{reject} = \frac{1}{n}\left(d_{reject} - \frac{1}{2}\right) \approx 0.059$$

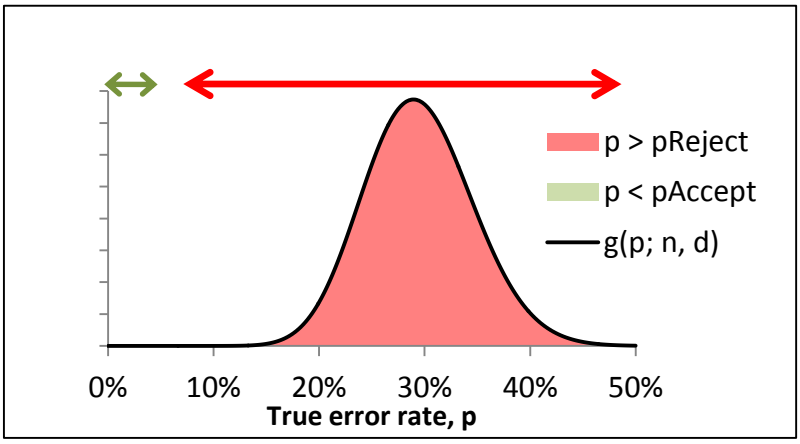

Figure 3. P_C = 0% and P_{NC} = 100% at d = 22 and n = 76 for batch ID 2783. Maximum point for the distribution is at p_{max} = 22/76 = 28.9%.

[6] D J Wheeler, D S Chambers, *Understanding Statistical Process Control*, 3rd edition, SPC Press

The rejection and acceptance limits are found in table (1) and $n = 76$. We calculate the probability that the true error rate is below the acceptable rate, $p \leq p_{Accept}$ (indicated by the green arrow in figure 3), and above the rejection rate, $p \geq p_{Reject}$ (indicated by the red arrow), by computing the area under the probability distribution:

$$P_C(\hat{p}) = \int_0^{p_{Accept}} g(p;\hat{p})dp \tag{5}$$

$$P_{NC}(\hat{p}) = \int_{p_{Reject}}^{1} g(p;\hat{p})dp \tag{6}$$

For the example batch the probability of conformance $P_C \approx 0$ and the probability of non-conformance $P_{NC} \approx 1$ because the observed rate of non-conformance is unusually high. Finally the risks are computed as the product of the probabilities and the costs:

$$\rho_{producer} = P_C \cdot 2000 \approx \underline{0}$$
$$\rho_{consumer} = P_{NC} \cdot 1064 \approx \underline{1064}$$

Here the average cost of rejection and replacement is NOK 2 000,- per unit, as explained previously. The accumulated cost of the measurement errors for this batch over 8 years continued operation is easily computed from the average annual cost error of NOK 133 (see figure 2) , and equals NOK 1 064. The conclusion is obvious in this case, we avoid the larger consumers risk by rejecting the batch and replacing it with new and improved devices.

Risk curves

The probabilities computed from equations (5) and (6) only depend on the sample size and the observed number of non-conforming instruments. It is therefore possible to compute P_C and P_{NC} for every possible observation for each of the sampling plans listed in table (1). Provided the test laboratories refrain from curtailment the computation can be done once and the actual decision only depends on the costs.

For small values of d the beta-distribution becomes asymmetric, even for sample sizes of the order of 100. Fig (4) shows how d and n affects the shape of the curve: the peak of the curve is always determined by $\hat{p} = d/n$, but for any given n it is asymmetric for small d, gradually becoming more symmetric as it increases.

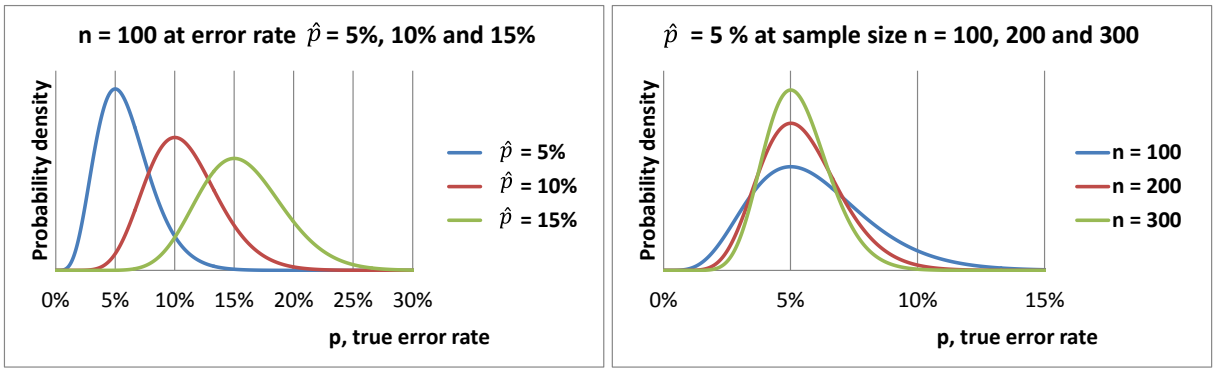

Figure 4: Probability density distributions for the true error rate, p, at different observations of the error rate, $\hat{p}$ and sample sizes, n.

Since small values of both $\hat{p}$ and n is often the case when assessing electrical energy meters, this non-symmetry is important when calculating probabilities in the tails of the distribution.

The values of p_{accept} and p_{reject} are different for the first and second part of the sampling plans. Fig (5) shows how, with the same observed error rate (4%), the first and second sampling plans produce different limiting values. In the first sampling plan there is a wide band with inconclusive test results. This band is closed in the second stage.

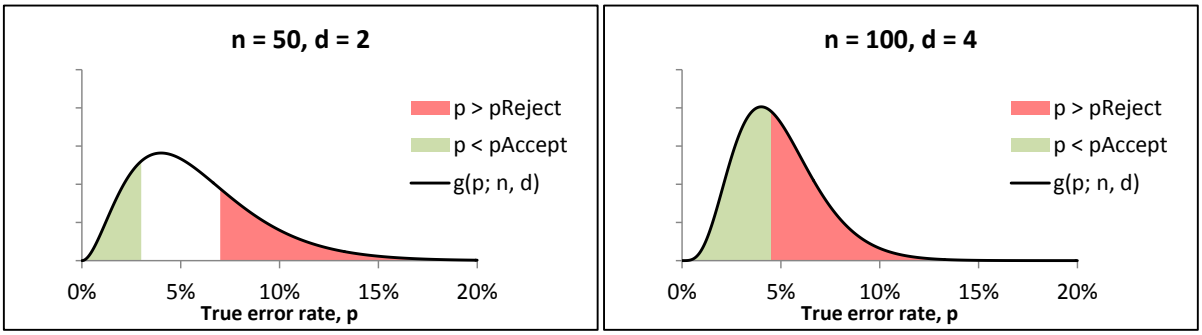

Figure 5. Beta-distribution for the error rate using attribute sampling, with indications of probabilities for conformance and non-conformance at first and second part of the sampling plan, row 2 in table 1.

In figure 6, P_C and P_{NC} is calculated for batch ID 2783 with n = 76 for different values of *d* from 0 up to 25.

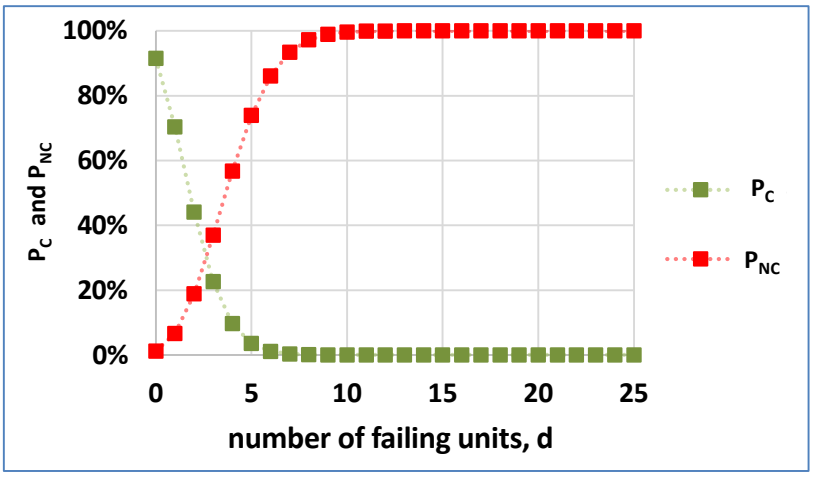

Figure 6. For batch id 2783, P_C and P_{NC} for different values of *d*.

To calculate the producers risk and the consumers risk, the probability curves in figure 6 are multiplied with the costs of making erroneous decisions, rejecting a conforming batch (producers risk) or accepting a non-conforming batch (consumers risk).

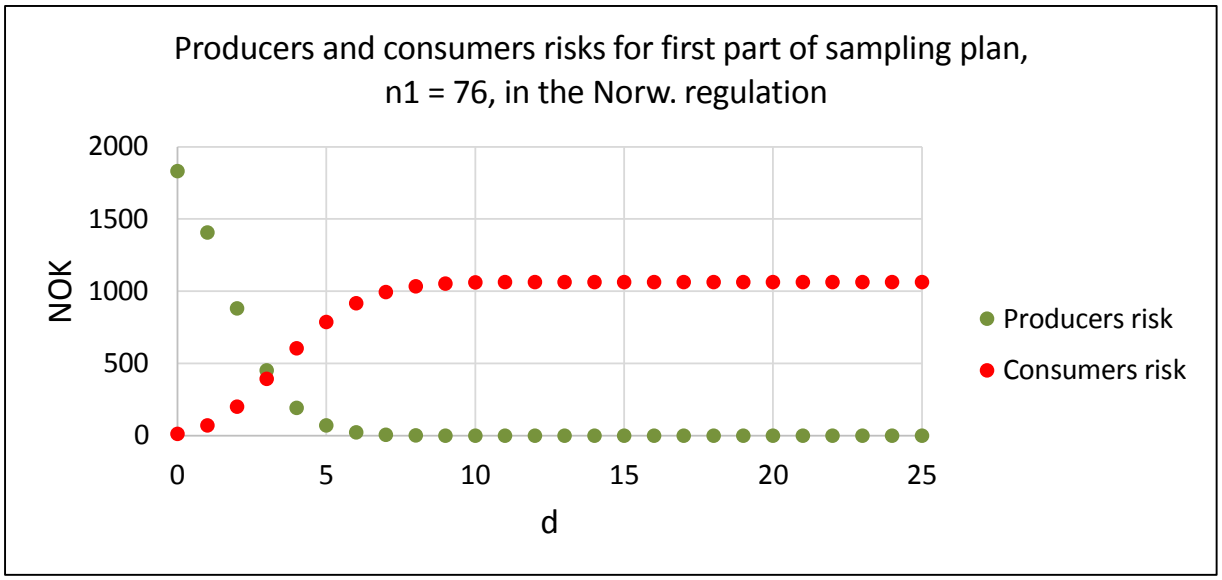

Figure 7 Average risk per instrument for producer (green dots) and consumer (red dots) for Batch id 2783, using watershed specifications and curtailment of the sampling.

In figure 7, consumer and producer risks are plotted as functions of *d* using a sampling plan defined in the Norwegian regulation. These risk curves display the risks as functions of the observed error rate,

$\hat{p} = d/n$, whereas risk operational characteristic curves[7] display the risks as a function of the true error rate, p. With $n_1 = 76$ and $d = 22$, the consumers risk is higher than the producers risk. We avoid the higher risk by rejecting the batch, and the producers risk of false rejection comes into play. If d was 0, 1, 2 or 3, the producers risk would be higher than the consumers risk. Then we could have avoided the higher risk by accepting the batch, and the consumers risk of false acceptance would have come into play.

If cost of replacement was higher, the producers risk curve would become higher and the crossing point between the risk curves would have occured at a higher value of d. If cost because of measurement errors was higher (due to larger measuring errors, higher price on the utility or a longer time until the next test) the consumers risk curve would have become higher, and the crossing point between the risk curves would have occured at a lower value of d.

Conformance assessment require high probability of conformance when we decide to accept the batch, and high probability of non-conformance when we reject the batch. We have seen that the costs involved with the two different decisions in a risk-based approach modify the criterion for when to accept and when to reject a batch. The modelling of the costs involved is therefore crucial to the risk analysis, and there are arguments against our simple cost calculations:

- Often the measuring error is negative, so the cost error is to the benefit of the consumer.

  - *MPE* in the regulation sets a limit to both positive and negative errors. It is unfair to all the other consumers that some have a utility meter with negative measuring errors. Since both negative cost errors and positive cost errors are equally important, we add them as absolute values, see Eqn (3).

- Consumers will also cover cost of replacement. Therefor a total cost calculation would be more adequate, and a method to minimize total costs would be more interesting for consumers.

  An important purpose of the regulation is to protect consumers against excessive measuring errors, and electric power is a kind of utility, which they are not capable of controlling themselves. The responsibility for the quality assessment of el-meters is placed on the grid owners (producer side), even though they might not have any special interest in a fair distribution of the costs. There are different ways to attend to this responsibility. Since installation cost is high compared to the cost of purchase, it is worthwhile to invest in accurate el-meters with a long life-time (in contrast to low cost, low quality el-meters with short lifetime), and they could follow up single el-meters showing suspicious behavior. Consumer or producer could easily become suspicious to the accuracy of the measurements if a unit has very large

[7] L.R. Pendrill 2008, "Operating 'cost' characteristics in sampling by variable and attribute" *Accred. Qual. Assur.*, **13**, 619 - 631, DOI: http://dx.doi.org/10.1007/s00769-008-0438-y

measuring errors. And the producer can offer a calibration of this unit and replace it if it does not comply with requirements. The batch might become accetable if we replace those el-meters, which are identified as outliers. However, this kind of work must be separated from the test regime of the regulation. If so, such work protects against false rejection of a batch during an acceptance sampling test. Then it is reasonable to accept a homogeneous batch with just a few failing units, because outliers could be detected by other means. And it is reasonable to reject a homogeneous batch with many failing units, because the entire batch is of bad quality according to the requirements.

If one considers only costs, the conclusion from our example test would have been to accept batch id 2783, because the cost pr unit of replacement is almost twice the cost of the measurement errors accumulated over 8 years of operation. Calculation of probability of conformance and probability of non-conformance and with no consideration to costs, indicate that we should reject this batch, figure 6. The combined effect of costs and probability of conformance / non-conformance in a risk analysis approach to decision-making, indicate that we should reject this batch, figure 7.