



ISTITUTO NAZIONALE DI RICERCA METROLOGICA Repository Istituzionale

Risks in a sausage conformity assessment due to measurement uncertainty, correlation and mass balance constraint

This is the author's submitted version of the contribution published as:

Original

Risks in a sausage conformity assessment due to measurement uncertainty, correlation and mass balance constraint / Pennechi, Francesca R.; Kuselman, Ilya; Di Rocco, Aglaia; Hibbert, D. Brynn; Semenova, Anastasia A. - In: FOOD CONTROL. - ISSN 0956-7135. - 125:(2021), p. 107949. [10.1016/j.foodcont.2021.107949]

Availability:

This version is available at: 11696/73034 since: 2022-02-17T10:00:54Z

Publisher:

ELSEVIER SCI LTD

Published

DOI:10.1016/j.foodcont.2021.107949

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

Journal of Food Composition and Analysis

Risks in a sausage conformity assessment due to measurement uncertainty, correlation and mass balance constraint

--Manuscript Draft--

Manuscript Number:	
Article Type:	Research Paper
Keywords:	Conformity assessment; Measurement uncertainty; Correlation; Mass balance constraint; False decision; Total risk; Multivariate Bayesian modelling; Monte Carlo method
Corresponding Author:	Ilya Kuselman, D.Sc. Independent Consultant on Metrology Modiin, ISRAEL
First Author:	Francesca R. Pennechi, PhD
Order of Authors:	Francesca R. Pennechi, PhD Ilya Kuselman, D.Sc. Aglia Di Rocco, M.Sc. David Brynn Hibbert, Em. Prof. Anastasia A. Semenova, Prof.
Abstract:	<p>A technique for evaluation of risks of false decisions in conformity assessment of the chemical composition of sausages was developed based on a multivariate Bayesian approach, taking into account measurement uncertainty, correlation and mass balance constraint of the regulated contents of the sausage components. As a case study, a dataset of test results of chemical compositions of the sausage "Braunschweigskaya" (measured contents of fat, protein, moisture and salt) was used for evaluations of the risks. A total global consumer's risk of 0.006 and a total global producer's risk of 0.017, characterizing the production process in general, were evaluated using Monte Carlo simulations. The difference in the risk values indicates a clear preference of the consumer's interests over the producer's interests. The total specific consumer's and producer's risks, related to a (specific) sausage batch, were evaluated using normal approximations in the Bayesian model. The risk values obtained were much more significant when measured contents approached their tolerance/specification limits or exceeded them. The codes, written in the R programming environment, for calculations of both the total global and the specific risks are provided as electronic supplementary material to this paper.</p>

Cover letter

Professor J. Stephen Elmore,

Editor-in Chief
Journal of Food Composition and Analysis

29 Nov, 2020

Dear Prof. Elmore,

Please find attached the manuscript by Francesca R. Pennechi, Ilya Kuselman, Aglaia Di Rocco, D. Brynn Hibbert, and Anastasia A. Semenova, titled "Risks in a sausage conformity assessment due to measurement uncertainty, correlation and mass balance constraint", which we would like to publish in Journal of Food Composition and Analysis.

Novelty Statement: A new technique for evaluation of the risks of false decisions in conformity assessment of chemical composition of sausages was developed based on the multivariate Bayesian approach and Monte Carlo simulations, taking into account the measurement uncertainty, correlation and mass balance constraint of the regulated sausage components contents. As a case study, a dataset of test results of chemical composition of sausage "Braunschweigs kaya" was used for evaluations of the risks.

Declaration of interest: There is no any actual or potential conflict of interest of any author.

Submission declaration and verification: The work described has not been published previously, it is not also under consideration for publication elsewhere. The publication in Journal of Food Composition and Analysis is approved by all the authors.

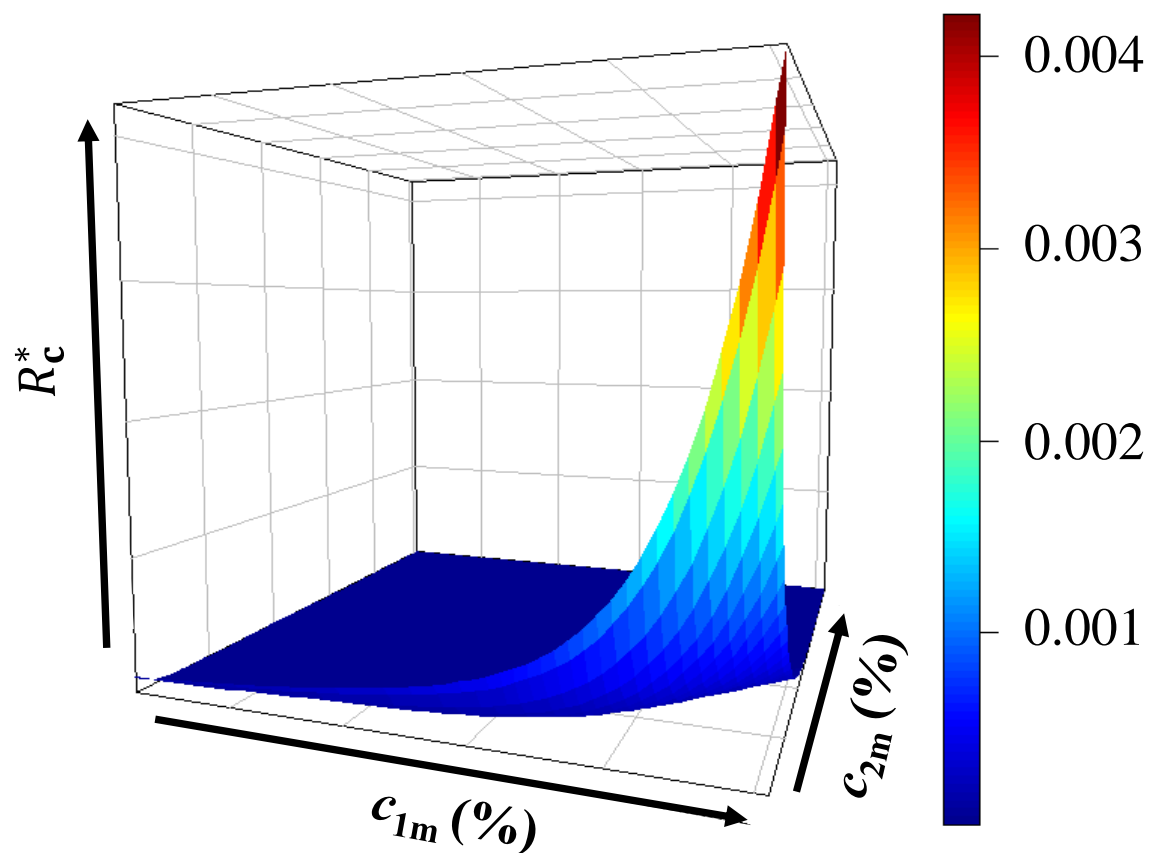
Best regards,



Dr. Ilya Kuselman,
the manuscript corresponding author,
ilya.kuselman@gmail.com
ilya.kuselman@bezeqint.net

HIGHLIGHTS

- A technique for evaluation of risks in a sausage conformity assessment is developed
- Measurement uncertainty, correlation and mass balance constraint are considered
- A multivariate Bayesian approach and a Monte Carlo method are applied
- Risks in assessment of sausage "Braunschweigs kaya" are evaluated as a case study



Surface of total specific consumer's risk R_c^* vs. measured contents of fat c_{1m} and protein c_{2m} in sausage "Braunschweigskaya".

Risks in a sausage conformity assessment due to measurement uncertainty, correlation and mass balance constraint

Francesca R. Pennechi^a, Ilya Kuselman^{b,*}, Aglaia Di Rocco^c, D. Brynn Hibbert^d,
and Anastasia A. Semenova^e

^a Istituto Nazionale di Ricerca Metrologica (INRIM), Strada delle Cacce 91, 10135 Turin, Italy

^b Independent Consultant on Metrology, 4/6 Yarehim St., 7176419 Modiin, Israel

^c University of Turin, Via Verdi 8, 10124, Turin, Italy

^d School of Chemistry, UNSW Sydney, Sydney NSW 2052, Australia

^e V.M. Gorbатов Federal Research Center for Food Systems of Russian Academy of Sciences,
26 Talalikhina St., 109316 Moscow, Russia

* Corresponding author. Tel.: +972-50-6240466

E-mail address: ilya.kuselman@bezeqint.net (I. Kuselman).

4/6 Yarehim St., Modiin, 7176419 Israel

ABSTRACT

A technique for evaluation of risks of false decisions in conformity assessment of the chemical composition of sausages was developed based on a multivariate Bayesian approach, taking into account measurement uncertainty, correlation and mass balance constraint of the regulated contents of the sausage components. As a case study, a dataset of test results of chemical compositions of the sausage "Braunschweigs kaya" (measured contents of fat, protein, moisture and salt) was used for evaluations of the risks. A total global consumer's risk of 0.006 and a total global producer's risk of 0.017, characterizing the production process in general, were evaluated using Monte Carlo simulations. The difference in the risk values indicates a clear preference of the consumer's interests over the producer's interests. The total specific consumer's and producer's risks, related to a (specific) sausage batch, were evaluated using normal approximations in the Bayesian model. The risk values obtained were much more significant when measured contents approached their tolerance/specification limits or exceeded them. The codes, written in the R programming environment, for calculations of both the total global and the specific risks are provided as electronic supplementary material to this paper.

Keywords:

Conformity assessment; Measurement uncertainty; Correlation; Mass balance constraint; False decision; Total risk; Multivariate Bayesian modelling; Monte Carlo method

1. Introduction

Many people like sausages and want to be sure that the chemical composition of a sausage bought in a store (a product of a known factory) corresponds to the requirements specified in national standards. Chemical composition determines the sausage taste, nutritional and energy value, and even safety, since such composition components as moisture and salt influence pathogenic microorganisms in the sausage. Of course, not only sausage consumers, but also producers are interested in an evidence that the chemical composition of a sausage batch put on sale meets the specified requirements.

Demonstration that the specified requirements are fulfilled is the result of product conformity assessment [1]. Standard specifications for chemical composition of a sausage, as a multicomponent material, are tolerance limits of the actual ('true') content c_i of the i -th component, $i = 1, 2, \dots, n$. Conformity assessment of an item (a sausage batch or sample) is based on comparing the measured content c_{im} with corresponding tolerance limits. Since any c_{im} value has an associated standard measurement uncertainty u_i [2, 3], acceptance limits for the measurement results can be used in addition to tolerance limits. In these cases, the decision rules (does the test item conform or not?) are based on comparing the measured content values c_{im} with the acceptance limits [4]. When tolerance limits have been defined by already taking into account measurement uncertainty, acceptance limits and tolerance limits coincide. Anyway, measurement uncertainty can cause several kinds of risk of a false decision on the conformity of an item.

The probability of accepting a sausage batch, when it should have been rejected, is the 'consumer's risk', whereas the probability of falsely rejecting a conforming batch is the 'producer's risk'. For a specified batch under control, they are referred to as the 'specific consumer's risk' and the 'specific producer's risk', respectively. The risks of incorrect conformity assessment of a batch randomly drawn from a statistical population of such batches are the 'global consumer's risk' and the 'global producer's risk', as they characterize the sausage production globally [4].

In general, a component-by-component evaluation of the risks of false decisions in a conformity assessment is not complete, as it does not give an answer to the question of the probability of a false decision on conformity of the product as a whole. When conformity assessment for each i -th component of a sausage batch is successful (i.e. the *particular* specific or global risks are small

enough), the total probability of a false decision concerning the batch as a whole (the *total* specific or *total* global risk) might still be significant. Evaluation of the total risks is detailed in IUPAC/CITAC Guide [5] based on a Bayesian multivariate approach.

Since the regulated components' contents of a sausage are subject to the mass balance constraint $\sum_i^n c_i = 100\%$, they are intrinsically correlated. This so-called 'spurious' correlation is observed in addition to other possible natural and/or technological correlations between the components' contents. All correlations may influence understanding of test results (c_{im} with associated u_i) and the evaluation of risks of false decisions in the conformity assessment. The circumstances mentioned above require appropriate modeling for the multivariate c_i distribution in different batches (prior distribution) and the multivariate c_{im} distribution in the same batch under test (likelihood function). An application of Monte Carlo simulations in the R programming environment can be helpful to overcome these difficulties [6].

The objective of the present paper is the development of a technique for evaluation of the risks in conformity assessment of chemical composition of sausages. As a case study, a dataset of compositions of sausage "Braunschweigs kaya", accumulated at V.M. Gorbатов Federal Research Center for Food Systems of Russian Academy of Sciences [7], is analyzed. The sausage name comes from the German Braunschweiger. It is a type of sausage prepared differently in different countries. In Russia "Braunschweigs kaya" is a summer (dry) sausage, produced since the end of the 19th century from a mixture of beef, pork and bacon with addition of salt and some other ingredients [8].

2. Material and test methods

Test results of the chemical composition of a total $N = 83$ batches of the sausage, produced according to the standardized technical conditions [9] during about three years at two similar factories, were studied as a case study of a dataset for quantification of the total risks. The testing was performed at the factory's laboratory for conformity assessment of a sausage batch to the Russian standard [10] before it is placed on the market. The dataset is provided as electronic supplementary material to this paper (RawData.txt file).

2.1. Tolerance and acceptance limits

The standard [9] sets the lower and upper tolerance/specification limits, T_{Li} and T_{Ui} , respectively, of contents c_i as mass fractions (%) of the four following chemical components in sausage "Braunschweigsкая":

$$i = 1) \text{ fat content } c_1 \leq 53.0 \% = T_{U1};$$

$$i = 2) \text{ protein content } c_2 \geq 15.0 \% = T_{L2};$$

$$i = 3) \text{ moisture content } c_3 \leq 40.0 \% = T_{U3};$$

$$i = 4) \text{ salt content } c_4 \leq 5.0 \% = T_{U4}.$$

Limitations of the sodium nitrite content (not exceeding 0.005 %) and pH (not less than 4.8) do not practically influence the mass balance $\sum_i^4 c_i = 100 \%$. Minor quantities of spices, nutritional supplements and starter cultures for fermentation, used in the process of the sausage preparation, are not regulated by the standard [9] and not related to the mass balance. They are not discussed further in the present work.

Acceptance limits, which take into account measurement uncertainties, are not applied in the factories, and measured values c_{im} are directly compared with the specification limits. Therefore, in the present study acceptance limits are taken as to be the same as specification limits.

2.2. Multivariate sub-domain of feasible sausage compositions

The specification limits of contents of the components, T_{Li} and T_{Ui} , form a multivariate specification domain of permissible sausage compositions. However, there is also the mass balance constraint which leads to a multivariate sub-domain of feasible compositions. For example, for fat content $c_1 = T_{U1} = 53.0 \%$, moisture content $c_3 = T_{U3} = 40.0 \%$ and salt content $c_4 = T_{U4} = 5.0 \%$, the protein content from the mass balance is $c_2 = 100 \% - (53.0 + 40.0 + 5.0) \% = 2.0 \%$, which is less than $T_{L2} = 15.0 \%$ and hence not permissible. On the other hand, a composition such as $c_1 = T_{U1}$, $c_2 = T_{L2}$, $c_3 = T_{U3}$ and $c_4 = T_{U4}$ is within the specification domain, but cannot be realized because it is in contradiction with the mass balance constraint.

Therefore, the multivariate sub-domain of feasible sausage compositions can be imagined as a part of the 3-simplex, which is the triangular pyramid shown in Fig. 1 by black lines, with c_1 , c_2 , c_3 and c_4 equal to 100 % in their vertices. The pyramid is truncated by the specification limits T_{U1} , T_{L2} , T_{U3} and T_{U4} , which are represented in Fig. 1 as the transparent triangular planes with colored

Fig. 1

borders. The facets of the resulting sub-domain of feasible compositions are highlighted with the colors of the corresponding specification limits.

2.3. Test methods and standard measurement uncertainties

A sample of a sausage of not less than 200 g is taken from a batch following the appropriate sampling standards [10, 11], homogenized and kept at 4 °C for not more than 24 hours, during which measurements of component contents should be performed.

The standard method [12] for measurement of a fat content c_1 is based on multiple fat extractions from the dried sample with a solvent (hexane, diethyl ether or petroleum ether) in a Soxhlet fat extraction apparatus. After that the solvent is removed and the fat dried to constant weight. In the interval of fat contents c_1 from 15 % to 50 % the maximum allowed relative error ($\pm \delta_1$) of the measured value c_{1m} at a level of confidence 0.95 and a normal distribution, set in the standard [12], is 10 %. Considering δ_1 as the target expanded measurement uncertainty [13] with coverage factor 2, the standard measurement uncertainty is $u_1 = (\delta_1/2 \cdot 100) c_{1m} = 0.05 c_{1m}$.

Protein content is measured by the standard Kjeldahl method [14]. Hot-acid digestion of a sample converts protein to ammonia, which is then distilled into a standardized acid, after which the acid is back-titrated and the result calculated. In the interval of protein contents c_2 from 20 % to 55 % the maximum allowed relative error of the measured value c_{2m} at the level of confidence 0.95 and a normal distribution is $\delta_2 = 8$ % according to the standard [14]. Hence, the standard measurement uncertainty is $u_2 = (\delta_2/2 \cdot 100) c_{2m} = 0.04 c_{2m}$.

The standard measurement method [15] for moisture content consists of drying a sample with sand to constant weight at a temperature of (103 ± 2) °C. In the interval of moisture contents c_3 from 1 % to 35 % the maximum allowed relative error of the measured value c_{3m} at a level of confidence 0.95 and a normal distribution is $\delta_3 = 12$ % [15]. Therefore, the standard measurement uncertainty is $u_3 = (\delta_3/2 \cdot 100) c_{3m} = 0.06 c_{3m}$.

Salt content c_4 is measured by Mohr's standard titration method [16]. This method determines chloride ions extracted from the sample by titration with silver nitrate. As the silver nitrate solution is slowly added, a precipitate of silver chloride forms. At the end point additional silver ions react with chromate ions of the indicator (potassium chromate) to form a red-brown precipitate of silver chromate. In the interval of salt contents c_4 from 3.5 % to 7.0 % the maximum allowed relative

error of the measured value c_{4m} at a level of confidence 0.95 and a normal distribution is $\delta_4 = 8\%$ [16]. The corresponding standard measurement uncertainty is $u_4 = (\delta_4/2 \cdot 100) c_{4m} = 0.04 c_{4m}$.

The standards [12,14,16] also include alternative test methods with similar metrological characteristics. Note that a metrologically-related correlation of the test results (between measured content values of different components of the same batch or sample) [6] is impossible here as the applied chemical analytical methods [12, 14-16] are based on different principles.

3. Modelling and calculation

3.1. Analysis of raw data

3.1.1. ANOVA

Analysis of variance (ANOVA) was performed to check homogeneity of the dataset consisting of two groups of test results of the sausage batches produced by factory A (the first 45 lines in the electronic supplementary material, RawData.txt file) and factory B (the last 38 lines in RawData.txt file). Empirical values F_i of the Fisher criterion in Table 1 are the ratios of the between-factories variance $s_{i\text{between}}^2$ to the within-factory variance $s_{i\text{within}}^2$. For fat, moisture and salt, i.e. for $i = 1, 3$ and 4 , F_i is less than the critical value $F_{\text{crit}} = 3.96$ for a significance level 0.05 and degrees of freedom of the between- and within- variances equal to $(2 - 1) = 1$ and $(45 + 38 - 2) = 81$, respectively [17]. As the significance level is the probability to reject the null hypothesis about equivalence of the variances when it is in fact true, $F_i < F_{\text{crit}}$ are indicating homogeneity of the data.

Table 1

The mean values of the protein content in the products of factory A and factory B are $m_{2A} = 25.0\%$ and $m_{2B} = 24.1\%$, respectively. They both are far from the lower tolerance limit $T_{L2} = 15.0\%$, and although the difference between them is statistically significant ($F_2 = 9.84$) it is negligible from a metrological point of view, being less than the standard measurement uncertainty $u_2 = 1.0\%$ of such protein contents. Therefore, all the data were used ($45 + 38 = 83$ test results) for further calculations being considered practically homogeneous.

3.1.2. Distributions of the test results

Distributions of the test results are characterized in Table 2, where $c_{i\min}$ and $c_{i\max}$ are the minimum and the maximum measured values of the i -th component content in the dataset, respectively; m_i and s_i are the mean and the standard deviation, respectively. The values of s_i are 1.5 to 2.5 times greater than the corresponding standard measurement uncertainties u_i of the corresponding mean component content, since they are influenced also by the variability of raw materials and conditions of the technological process during the three years of sausage production, when the dataset was accumulated.

Table 2

Histograms of the distributions are presented in Fig. 2. Probability density functions (pdfs) of the fitting normal (theoretical) distributions are shown in Fig. 2 as solid curves. Goodness-of-fit was evaluated by the one-sample Kolmogorov-Smirnov test [18]. Calculated values of the test statistic D_i and corresponding probabilities P_i are shown in Table 2. As $P_i > 0.05$, the hypothesis on goodness-of-fit is not rejected, i.e. the data distribution does not differ significantly from normal.

Fig. 2

Note, also the protein content distribution is well fitted by a normal distribution, as seen by the values of D_2 and P_2 , thereby supporting the decision made in Sec. 3.1.1 on the practical homogeneity of the corresponding data.

3.1.3. Correlations and covariances

Calculated Pearson's correlation coefficients r_{ij} ($i \neq j$, $j = 1, 2, 3, 4$) between measured components' contents are presented in Table 3. The critical two-tailed value for $(83 - 2) = 81$ degrees of freedom and the level of confidence 0.95 is 0.216, i.e. r_{ij} is considered not significant when $|r_{ij}| \leq 0.216$ [17].

Table 3

The greatest absolute value of the correlation coefficient is related to contents of fat and moisture, $r_{13} = r_{31} = -0.318$. It is a negative correlation caused by the mass balance constraint: when fat content increases it is mainly at the expense of the moisture content. A similar negative correlation is seen between fat and salt contents, as well as between protein and moisture contents. Correlation between fat and protein contents is also negative, but statistically negligible, as $|r_{12}| = |r_{21}| = |-0.163| < 0.216$. The reason is that both fat and protein enter the sausage with meat: beef contains on average about 4 % of fat, pork 14 %, and bacon 93 %. To a certain extent this might

have contributed a positive correlation, softening the original negative correlation due to the mass balance constraint. There is only one positive correlation coefficient in Table 3, $r_{24} = r_{42} = 0.301$, which characterizes the relatively strong dependence between protein and salt contents, probably arising when the minced meat is dried.

The covariance matrix presented in Table 4 shows variances s_i^2 (s_i are in Table 2) as diagonal elements, and covariances $cov_{ij} = r_{ij} s_i s_j$, $i \neq j$ (r_{ij} are in Table 3) as the off-diagonal elements.

Table 4

Note that the distributions of the components' contents shown in Fig. 2 are the marginal distributions of a multivariate pdf, whose covariance matrix is that in Table 4. Such a multivariate pdf is used in the following for modelling the sausage compositions and the test/measurement results.

3.2. Bayesian modelling of total risks in conformity assessment

Bayes' theorem for the multivariate pdf of n components' contents in a sausage, as a multicomponent material, is expressed by the following equation:

$$g(\mathbf{c} | \mathbf{c}_m) = C g_0(\mathbf{c}) h(\mathbf{c}_m | \mathbf{c}), \quad (1)$$

where $\mathbf{c} = [c_1, c_2, \dots, c_n]$ and $\mathbf{c}_m = [c_{1m}, c_{2m}, \dots, c_{nm}]$ are vectors of the actual ("true") values c_i and measured values c_{im} , respectively, $i = 1, 2, \dots, n$; $g(\mathbf{c} | \mathbf{c}_m)$ is the multivariate (joint) posterior pdf; C is a normalizing constant; $g_0(\mathbf{c})$ is the multivariate prior pdf taking into account correlations between c_i ; and $h(\mathbf{c}_m | \mathbf{c})$ is the multivariate likelihood function involving the measurement uncertainties and correlations between c_{im} [5].

The total global consumer's risk R_c and producer's risk R_p are, respectively [6]:

$$R_c = \int_{T^c} \int_A g_0(\mathbf{c}) h(\mathbf{c}_m | \mathbf{c}) d\mathbf{c}_m d\mathbf{c} \quad \text{and} \quad R_p = \int_T \int_{A^c} g_0(\mathbf{c}) h(\mathbf{c}_m | \mathbf{c}) d\mathbf{c}_m d\mathbf{c}, \quad (2)$$

where T is the tolerance/specification domain $T_1 \times T_2 \times \dots \times T_n$, A is the acceptance domain $A_1 \times A_2 \times \dots \times A_n$, and the integral symbols indicate multiple integrals. Superscript "c" of T in the formula for R_c means "complementary" for at least one T_i , whereas the integration with respect to all c_{im} is

performed within A . The subscript “c” of A in the formula for R_p means “complementary” for at least one A_i , whereas the integration with respect to all c_i is performed within T .

The total specific consumer’s risk R_c^* and total producer’s risk R_p^* are, respectively [19]:

$$R_c^* = 1 - \int_T g(\mathbf{c} | \mathbf{c}_m) d\mathbf{c} \text{ when } \mathbf{c}_m \text{ is in } A, \text{ and}$$

$$R_p^* = \int_{T_1} \dots \int_{T_\nu} \int_0^{100} \dots \int_0^{100} g(\mathbf{c} | \mathbf{c}_m) d\mathbf{c} \text{ when } c_{im}, 1 \leq i \leq \nu, \text{ are outside } A. \quad (3)$$

Here, R_c^* is the probability that at least one of corresponding true content values c_i of a sausage components is actually outside its tolerance interval T_i , when all the measured content values c_{im} are in their acceptance intervals A_i (false conforming). Thus, the R_c^* value equals to one minus the probability that all c_i are within tolerance domain T at the condition that all c_{im} conform, i.e. are within acceptance domain A .

Symbol ν in Eq. (3) for R_p^* indicates the number of those sausage components, $1 \leq \nu \leq n$, whose measured content values c_{im} are outside their acceptance intervals A_i . Hence, vector \mathbf{c}_m , being out of the acceptance domain A for those ν components of the tested product (a sausage batch), is rejected as non-conforming. For simplicity, and without losing generality, the measured values c_{im} outside their acceptance intervals are the first ν . Given that these ν measured values do not conform, R_p^* is the probability that the corresponding true values are all actually inside their tolerance intervals, hence the sausage batch satisfies its specifications and rejection of the batch is a false decision.

3.2.1. Multivariate prior pdf

The raw data discussed in Sec. 3.1 were used for modelling the multivariate prior pdf, which is a "theoretical" pdf of the actual/true values $\mathbf{c} = [c_1, c_2, c_3, c_4]$, based on the best available knowledge.

A total $M = 1 \cdot 10^7$ Monte Carlo (MC) simulations of the actual sausage compositions $\mathbf{c} = [c_1, c_2, c_3, c_4]$ were performed using a multivariate truncated normal pdf, according to Model 1 in paper

[6], having as location parameter the mean vector $\boldsymbol{\mu} = [m_1, \dots, m_4]$ (m_i are in Table 2), and as scale parameter the covariance matrix in Table 4.

True values have no uncertainties by definition [20] and their sum is to be exactly equal to 100 %, according to the law of conservation of mass [21]. Therefore, data drawn from the multivariate normal pdf truncated on the domain $[0, 100]^4$ should also be subjected to the closure operation [6]:

$$clo(\mathbf{c}) = \left[\frac{100 c_1}{\sum_{i=1}^4 c_i}, \dots, \frac{100 c_4}{\sum_{i=1}^4 c_i} \right]. \quad (4)$$

The resulting correlation matrix is given in Table 5. Comparing the correlation matrices in Table 3 and Table 5, one can see that the correlation coefficient related to contents of fat and moisture is still negative, but its absolute value is much larger. Also, the positive correlation coefficient between protein and salt contents is increased due to the closure operation. The corresponding covariance matrix is shown in Table 6.

The probability of conformance of the multivariate prior pdf, calculated as the fraction of M of the events when the obtained (simulated and closed) sausage compositions $\mathbf{c} = [c_1, c_2, c_3, c_4]$ were within the tolerance domain T , was $P_{\text{conf}} = 0.972$.

Note that random generation from a multivariate normal pdf (and subsequent closure of the generated data), without truncation on the domain $[0, 100]^4$, leads here to the same conformance probability $P_{\text{conf}} = 0.972$. This coincidence is caused by the distance of the contents of each component in the raw data (Table 2) from the truncation limits 0 % and 100 %, seen also in the histograms of Fig. 2.

3.2.2. Multivariate likelihood function

Modelling of the multivariate likelihood function for measured values $\mathbf{c}_m = [c_{1m}, c_{2m}, c_{3m}, c_{4m}]$ is based on the idea that a multivariate truncated normal pdf with zero expectation can model an error $\mathbf{e}_m = [e_{1m}, e_{2m}, e_{3m}, e_{4m}]$ which is then translated to the vector of actual content values $\mathbf{c} = [c_1, c_2, c_3, c_4]$ generated from the multivariate prior pdf. Therefore, \mathbf{c}_m is recovered as $\mathbf{c}_m = \mathbf{c} + \mathbf{e}_m$ [6].

The covariance matrix, used as the scale parameter of the truncated pdf associated with vector \mathbf{e}_m , is given in Table 7. The diagonal elements of this matrix are squared measurement uncertainties u_i^2 , obtained from u_i discussed in Sec. 2.3, and the off-diagonal elements are covariance terms equal to products $cov_{ijlf} = r_{ij} u_i u_j$ whose correlation coefficients r_{ij} are in Table 3. Subscript ‘lf’ in cov_{ijlf} means ‘likelihood function’.

Thus, modelling of both the likelihood function and the prior pdf is based on the experimental data, containing the initial knowledge about the sausage compositions. However, the likelihood function characterizes the measurement process with corresponding measurement uncertainties, mimicked by $\mathbf{e}_m = [e_{1m}, e_{2m}, e_{3m}, e_{4m}]$. Therefore, the measured values $\mathbf{c}_m = [c_{1m}, c_{2m}, c_{3m}, c_{4m}]$ are no longer required to sum to 100 %, and so the closure operation is not applied here.

3.2.3. Posterior multivariate pdf

Once the multivariate prior pdf and likelihood function are modelled, the posterior multivariate/joint pdf of the actual content values \mathbf{c} in a sausage at the measured values of the components’ contents \mathbf{c}_m can be calculated by Eq. (1) as the normalization of the product $g_0(\mathbf{c})h(\mathbf{c}_m | \mathbf{c})$.

The posterior pdf contains an updated state of knowledge about the product. Since this pdf takes into account what may happen during the measurement process, as with the likelihood data, the closure operation is not appropriate for the posterior data.

4. Results and discussion

4.1. Global risks

The total global consumer’s risk $R_c = 0.006$ was numerically recovered, according to Eq. (2), as the fraction of the M generated vectors of true and measured values $[c_1, c_2, c_3, c_4, c_{1m}, c_{2m}, c_{3m}, c_{4m}]$ in which all the measured values c_{im} were within their A_i but at least one of the corresponding true value c_i was outside T_i .

Correspondingly, the total global producer's risk $R_p = 0.017$ was evaluated as the fraction of the M generated vectors $[c_1, c_2, c_3, c_4, c_{1m}, c_{2m}, c_{3m}, c_{4m}]$ in which all the true values c_i were within their T_i , while at least one of the corresponding measured value c_i was outside A_i .

That means that six only from a thousand sausage batches may be falsely assessed as corresponding to the specifications, whereas seventeen conforming batches have a chance to be falsely rejected. This indicates a clear preference of the consumer's interests over the producer's interests. However, the discussed risks are only probabilities, and do not take into account the severity of the risks [5, 6]. Obviously, a producer is interested in maintaining the satisfaction of his consumers no less than the consumers themselves.

The code, written in the R programming environment [22], for calculations of the total global risks is provided as electronic supplementary material to this paper (R_code.r file).

4.2. Specific risks

For each specified vector of measured values $[c_{1m}, c_{2m}, c_{3m}, c_{4m}]$, the integrals of the posterior pdf in Eqs. (3) involve multiple integrals of the joint pdf of vector $[c_1, c_2, c_3, c_4, c_{1m}, c_{2m}, c_{3m}, c_{4m}]$ with respect to variables c_i ($i=1, \dots, 4$) over appropriate domains. $M = 10^7$ random vectors $[c_1, c_2, c_3, c_4, c_{1m}, c_{2m}, c_{3m}, c_{4m}]$ were generated, according to the prior modelling for c_i values in Sec. 3.2.1 and the likelihood modelling for c_{im} values in Sec. 3.2.2, as detailed in the paper [19, Sec. 4.4.2].

The results of calculations of the total specific consumer's risk R_c^* were practically zero (less than 0.001) at the vector of measured values \mathbf{c}_m containing c_{im} equal to the prior means m_i of the distributions in Table 2. When c_{im} moves away from m_i toward the tolerance limits, the R_c^* values are naturally increasing. The mean m_1 of the fat contents is far from their tolerance limit T_{U1} for more than three standard deviations s_1 in Table 2. Therefore, the influence of c_{1m} on R_c^* is minor. As the distance of the prior mean m_2 of the protein contents from their tolerance limit T_{L2} is greater than five standard deviations s_2 , the c_{2m} influence on R_c^* is also very minor. The distances of the mean m_3 of the moisture content from the tolerance limit T_{U3} , and of the mean m_4 of the salt content from the tolerance limit T_{U4} , are less than three standard deviations s_3 and s_4 , respectively, and as a result have an influence on the risk. For example, $R_c^* = 0.039$ at the measured fat content $c_{1m} = m_1 = 40.5\%$, protein content $c_{2m} = m_2 = 24.6\%$, moisture content $c_{3m} = 35.7\%$ (about 1.5 standard

deviations from m_3 towards the tolerance limit T_{U3}) and salt content $c_{4m} = 4.79\%$ (about 2 standard deviations from m_4 towards the tolerance limit T_{U4}).

Note that even for $M = 10^7$ simulations there were certain instabilities (and maybe bias) in the risk values, increasing when working in the tails of the posterior distribution, where only a few random numbers could be generated [23, 24]. Therefore, in the considered case, the MC method proved to be less reliable for calculation of risks than an analytical approximation method based on normal distributions [5].

Since the marginal distributions of the available data were successfully approximated by normal distributions (Sec. 3.1.2) and there was no evidence of an effect of truncation in the corresponding marginal distributions of the joint prior pdf, this pdf was constructed as a multivariate normal function with mean (Table 2) and covariance matrix (Table 6) estimated from the random values generated as described in Sec. 3.2.1. Because of the negligible effect of truncation, the multivariate likelihood function described in Sec. 3.2.2 was also approximated by a multivariate normal function having the covariance matrix shown in Table 7. Resorting to such approximations for the prior pdf and the likelihood function, a multivariate normal posterior pdf with parameters calculated as in IUPAC/CITAC Guide [5, Eq. (34)] was applied for R_c^* and R_p^* calculations according to Eqs. (3). The code for calculations of total specific risks is provided as electronic supplementary material to this paper, (R_code.r file).

The dependence of total specific consumer's risk on measured values of fat and protein contents is shown in Fig. 3a for cases in which measured contents of moisture and salt are constant and equal to their prior means. The R_c^* values are practically zero at the majority of combinations of c_{1m} and c_{2m} on the intervals from their prior means to the tolerance limits, and increase to 0.004 only, if simultaneously c_{1m} and c_{2m} are equal to their tolerance limits T_{U1} and T_{L2} , a case not observed in the raw data described in Table 2. In other words, the influence of fat content and protein content in the sausage on the total specific consumer's risk is negligible, as already explained above. The most influential variables here are contents of moisture and salt. The dependence of R_c^* on c_{3m} and c_{4m} on the intervals from their prior means to the tolerance limits is illustrated in Fig. 3b for cases in which measured contents of fat and protein are constant and equal to their prior means. The range of R_c^* values in this plot spreads from practically zero to 0.34. Of these two variables, the measured salt content has the most influence. Again, the risk is greatest when both c_{3m} and c_{4m} are at their tolerance limits.

Fig. 3

Fig. 4

The dependence of total specific producer's risk on measured values of moisture and salt contents is shown in Fig. 4 for cases in which the measured contents of fat and protein are constant and equal to their prior means, as in Fig. 3b. The interval of the measured moisture contents in Fig. 4a is from the tolerance limit T_{U3} to $1.18 T_{U3}$, i.e. to the three standard measurement uncertainties ($3 \times 0.06 T_{U3}$), while the interval of the measured salt contents is from the prior mean to the tolerance limit. Hence, in this particular case, measured contents of one component only do not conform. The calculated R_p^* values are from 0.69 to 1. Fig. 4b is the other way round with respect to Fig. 4a: here the measured salt contents do not conform, being in the interval from the tolerance limit T_{U4} to $1.12 T_{U4}$ (the three measurement standard uncertainties being $3 \times 0.04 T_{U4}$), while the interval of the measured moisture contents is from the prior mean to its tolerance limit. The R_p^* values go from practically zero at the $c_{4m} = 1.12 T_{U4}$, up to 0.77 when $c_{3m} = m_3$ and simultaneously $c_{4m} = T_{U4}$. The logic is that there is no producer's risk, if a component content in the tested batch substantially exceeded its tolerance limit and the batch is rejected. On the other hand, R_p^* can be extremely high when a measured content value is close to its tolerance limit. Correlations complicate the picture, but do not change it in essence.

In general, although the total global consumer's risk $R_c = 0.006$ and producer's risk $R_c = 0.017$ are small, the total specific risks related to a (specific) sausage batch can be much more significant when measured contents of moisture or salt, or both, are close to their tolerance limits or exceed them.

5. Conclusions

A technique for evaluation of the risks of false decisions in conformity assessment of chemical composition of sausages was developed based on the multivariate Bayesian approach, taking into account the mass balance constraint of the regulated contents of the sausage components. As a case study, a dataset of test results of the chemical composition of the sausage "Braunschweigs kaya" was used for evaluation of the risks. The marginal distributions of contents of the main (regulated) components – fat, protein, moisture and salt – could be approximated by normal probability density functions (pdfs). However, since these components contents are subject of the mass balance and limited on the [0 %, 100 %] domain, the prior multivariate distribution

describing the dataset was numerically simulated starting from a multivariate truncated normal pdf and the subsequent closure of the sum of the generated contents to 100 %. Measured values were modelled by a multivariate truncated normal likelihood function taking into account associated measurement uncertainties, and therefore the closure operation did not apply.

Based on these Monte Carlo simulations of the product of the prior pdf and the likelihood function, a total global consumer's risk of 0.006 and a total global producer's risk of 0.017 were evaluated, characterizing the production process in general. The difference in the risks' values indicates a clear preference of the consumer's interests over the producer's interests.

The Monte Carlo simulations were not sufficiently stable and reliable when working in the tails of the posterior pdf. Therefore, in order to calculate the total specific risks related to a (specific) sausage batch, an approximation for both the prior pdf and the likelihood function with a multivariate normal distribution (allowed by a negligible truncation effect in the considered case) was made in order to deal with a "plain" multivariate normal distribution for the posterior pdf. The total specific consumer's risk and producer's risk were shown to be significant when measured contents of moisture or salt were close to their tolerance (specification) limits or exceeded them. The contents of fat and protein did not practically influence these risks, being far from the specification limits.

CRedit authorship contribution statement

Francesca R. Pennechi: Methodology, Formal analysis, Software, Visualization, Writing – Reviewing and editing; **Ilya Kuselman:** Conceptualization, Project administration, Visualization, Writing - Original draft preparation; **Aglaia Di Rocco:** Software, Data curation, Visualization; **D. Brynn Hibbert:** Writing – Reviewing and editing; **Anastasia A. Semenova:** Resources, Validation.

Declaration of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research was supported in part by the International Union of Pure and Applied Chemistry (Project 2019-012-1-500), as well as by the Ministry of Science and Higher Education of the Russian Federation (Project 075-15-2020-775).

References

- [1] ISO/IEC 17000 (2004). Conformity assessment. Vocabulary and general principles.
- [2] JCGM 100 (2008). Evaluation of measurement data – Guide to the Expression of Uncertainty in Measurement (GUM). <https://www.bipm.org/en/publications/guides/>. Accessed 4 Aug 2020.
- [3] Ellison S.L.R. and Williams A. (Eds.) (2012). Eurachem/CITAC Guide: Quantifying uncertainty in analytical measurement. <https://www.eurachem.org/index.php/publications/guides/>. Accessed 4 Aug 2020.
- [4] JCGM 106 (2012). Evaluation of measurement data – The role of measurement uncertainty in conformity assessment. <http://www.bipm.org/en/publications/guides/>. Accessed 4 Aug 2020.
- [5] Kuselman, I., Pennechi, F.R., da Silva, R.J.N.B., Hibbert, D.B. (2020). IUPAC/CITAC Guide: Evaluation of risks of false decisions in conformity assessment of a multicomponent material or object (IUPAC Technical Report). *Pure Appl Chem*. <http://dx.doi.org/10.1515/pac-2019-0906>.
- [6] Pennechi F.R., Di Rocco, A., Kuselman, I., Hibbert, D.B., Sega, M. (2020). Correlation of test results and influence of a mass balance constraint on risks in conformity assessment of a substance or material. *Measurement*. <http://dx.doi.org/10.1016/j.measurement.2020.107947>.
- [7] V.M. Gorbatov Federal Research Center for Food Systems of Russian Academy of Sciences, Moscow, Russia. <http://www.vniimp.ru/>. Accessed 3 Aug 2020.
- [8] Semenova, A.A., Nasonova, V.V., Mileenkova, E.B. (2015). In best traditions of meat business. *All About Meat*, 3, 52-56 (in Russian). <https://elibrary.ru/item.asp?id=23710695>. Accessed 26 Aug 2020.

- [9] GOST 55456-2013 (2014). Dry sausages - Technical conditions. Standardinform, Moscow (in Russian).
- [10] GOST 9792-73 (2009). Sausage products and products of pork, mutton, beef and meat of other kinds of slaughter animals and poultry - Acceptance rules and sampling methods. Standardinform, Moscow (in Russian).
- [11] GOST 7269-79 (2006). Meat – Methods of sampling and organoleptic methods of freshness test. Standardinform, Moscow (in Russian).
- [12] GOST 23042-2015 (2016). Meat and meat products – Methods of fat determination. Standardinform, Moscow (in Russian).
- [13] Bettencourt da Silva, R., Williams A. (Eds) (2015). Eurachem/CITAC Guide: Setting and using target uncertainty in chemical measurement. <https://www.eurachem.org/index.php/publications/guides/>. Accessed 11 June 2020.
- [14] GOST 25011-2017 (2017). Meat and meat products – Protein determination methods. Standardinform, Moscow (in Russian).
- [15] GOST 33319-2015 (2016). Meat and meat products – Method for determination of moisture content. Standardinform, Moscow (in Russian).
- [16] GOST 9957-2016 (2016). Meat and meat products - Methods for determination of sodium chloride. Standardinform, Moscow (in Russian).
- [17] NIST/SEMATECH. e-Handbook of statistical methods. <https://www.itl.nist.gov/div898/handbook/pmd/section1/pmd144.htm>. Accessed 15 July 2020.
- [18] D’Agostino, R.B., Stephens, M.A. (Eds) (1986). Goodness-of-fit techniques, Statistics: a series of textbooks and monographs. New York, USA: Marcel Dekker Inc.
- [19] Pennechi, F.R., Kuselman, I., Di Rocco, A., Hibbert, D.B., Sobina, A., Sobina, E. (2020). Specific risks of false decisions in conformity assessment of a substance or material with a mass balance constraint – A case study of potassium iodate. *Measurement*. <https://doi.org/10.1016/j.measurement.2020.108662>.
- [20] JCGM 200 (2012), International Vocabulary of Metrology – Basic and General Concepts and Associated Terms. <http://www.bipm.org/en/publications/guides/>. Accessed 15 Aug 2020.
- [21] Tamir, A. (2013). Conservation law of mass. *J Chem Eng Process Technol*, 4, e114. <http://dx.doi.org/10.4172/2157-7048.1000e114>.

- [22] The R project for statistical computing. <https://www.r-project.org/>. Accessed 15 Aug 2020.
- [23] EPA/630/R-97/001 (1997). Guiding principles for Monte Carlo analysis. <https://www.epa.gov/risk/guiding-principles-monte-carlo-analysis>. Accessed 15 Nov 2020.
- [24] Botev, Z., L'Ecuyer, P. (2017) Simulations from the normal distribution truncated to an interval in the tail. In *Proceedings of the 10th EAI International Conference on Performance Evaluation Methodologies and Tools*. <http://dx.doi.org/10.4108/eai.25-10-2016.2266879>.

Figure 1

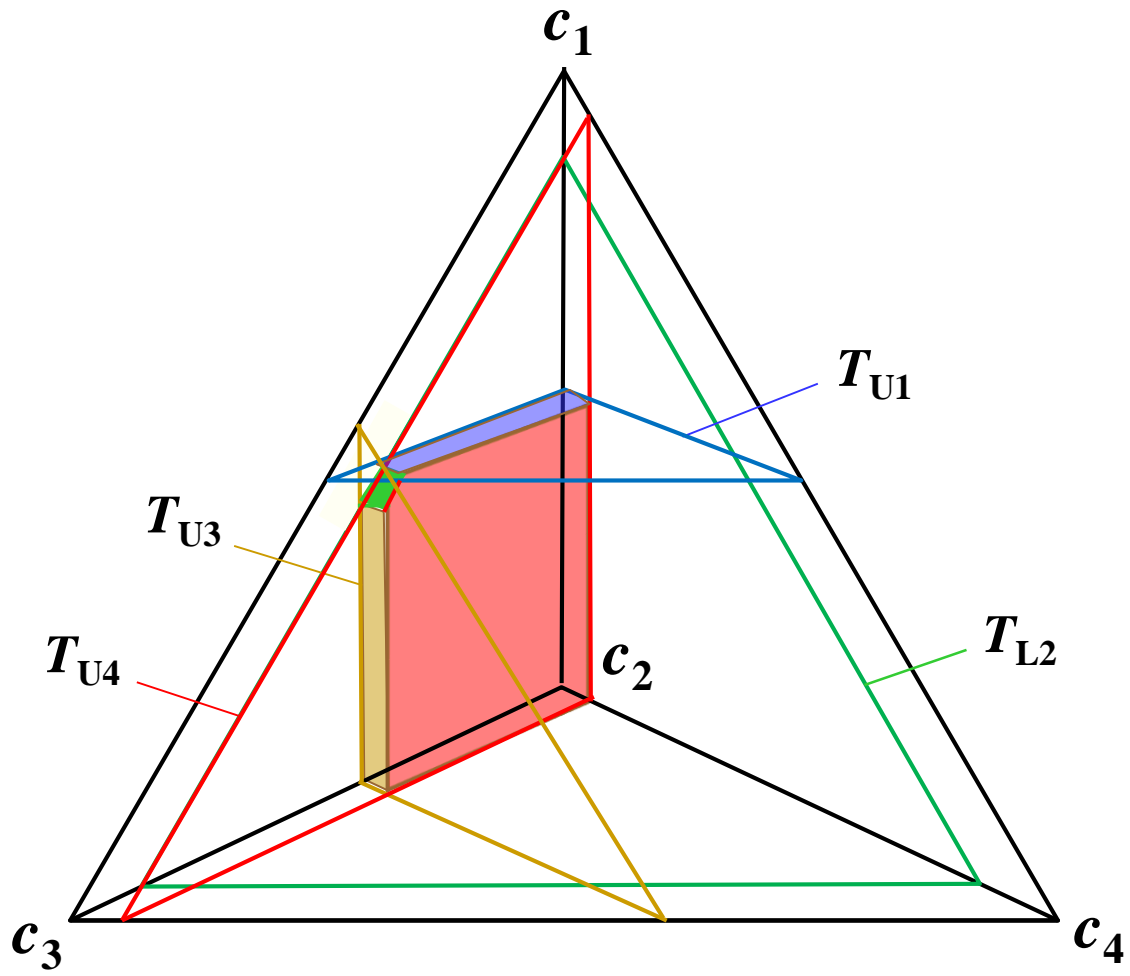


Figure 2

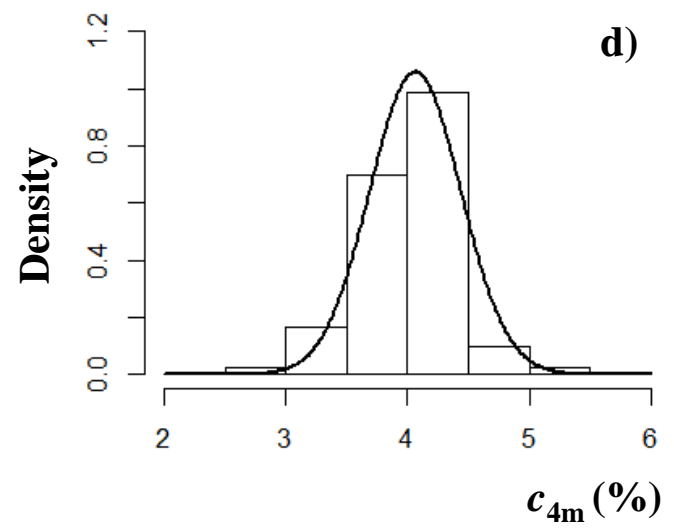
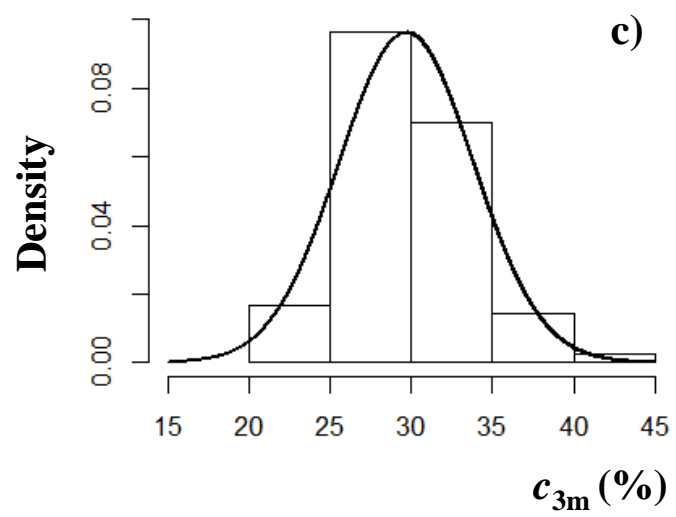
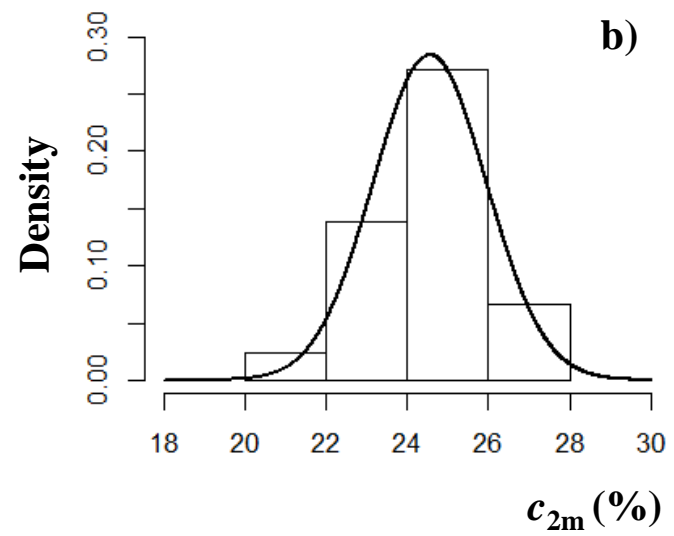
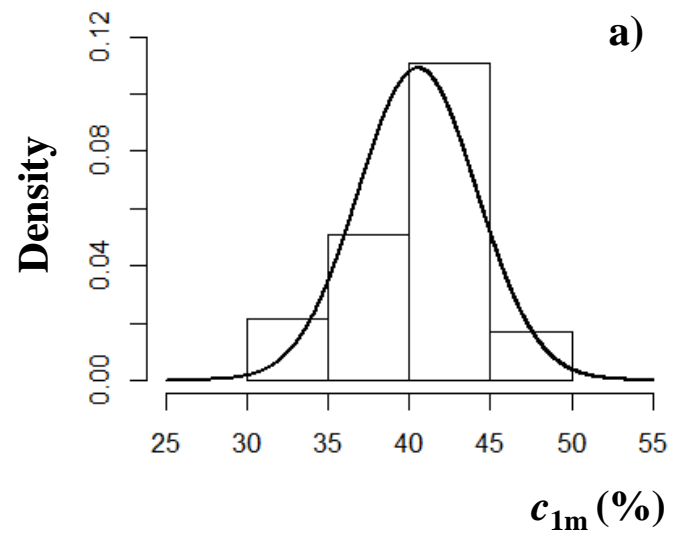


Figure 3

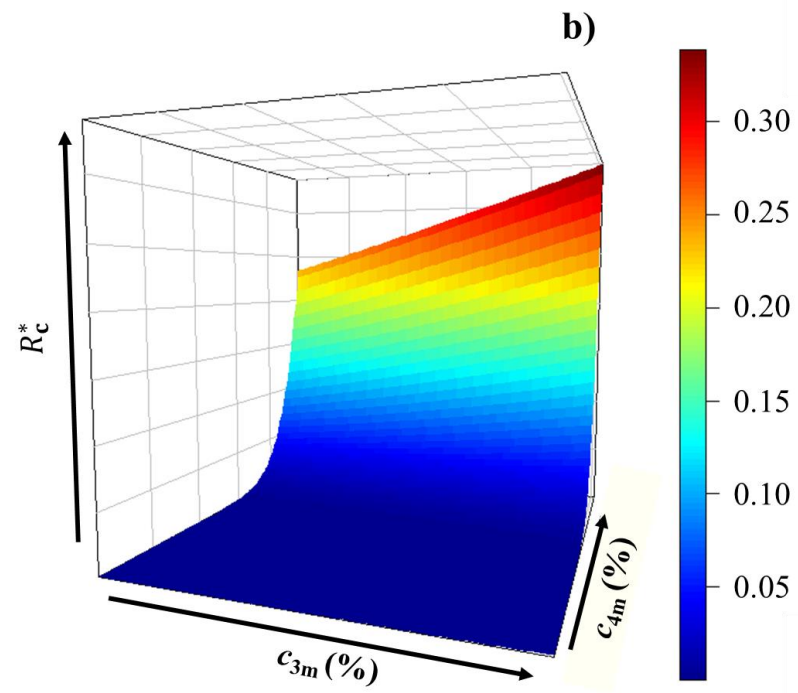
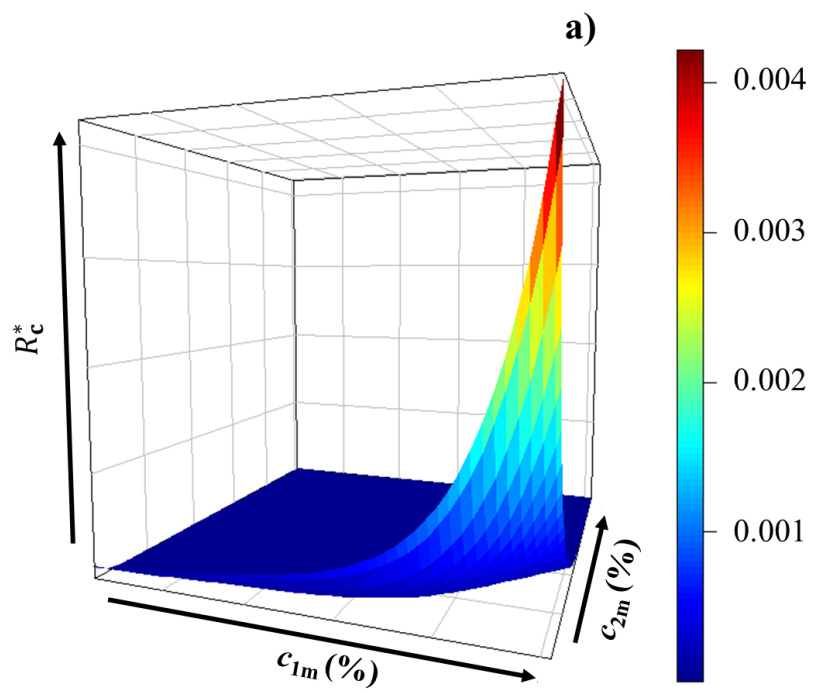


Figure 4

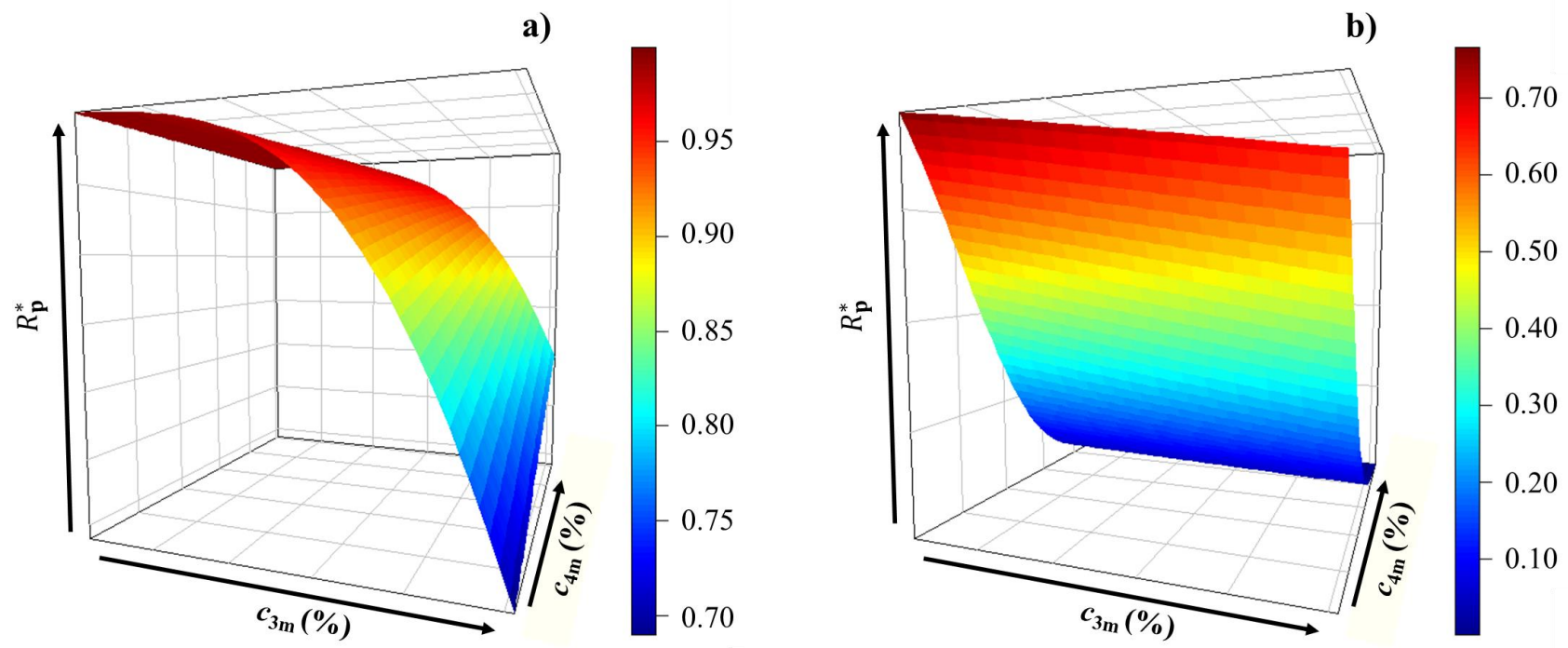


Figure captions

Fig. 1. Triangular pyramid of the sausage compositions. Each vertex corresponds to a component content c_i equal to 100 %. The permissible compositions are the part of the pyramid delimited by the tolerance/specification limits T_{U1} (the transparent triangle shown by blue border lines), T_{L2} (green lines), T_{U3} (brown lines) and T_{U4} (red lines). The left green border of T_{L2} is hidden by the red border of T_{U4} . The facets of the sub-domain of feasible compositions are marked with the colors of the tolerance limits. The left upper corner of the sub-domain is cut by the T_{L2} plane.

Fig. 2. Histograms of distributions of the measured components' contents and pdfs of fitted normal distributions. Plots a) – d) are for contents of fat c_{1m} , protein c_{2m} , moisture c_{3m} , and salt c_{4m} , respectively.

Fig. 3. Surface of R_c^* vs. measured values of the contents of the sausage components. The plot in Fig. 3a shows total specific consumer's risk R_c^* in dependence on measured contents of fat c_{1m} and protein c_{2m} from their prior means to the tolerance limits, while the measured moisture content and salt content are constant and equal to their prior means. The second plot, in Fig. 3b, demonstrates the R_c^* surface for the cases when the measured contents of fat and protein are constant and equal to their prior means, while the contents of moisture c_{3m} and salt c_{4m} are from their prior means to the tolerance limits. A color column bar gives indication of the risk values between the minimum and the maximum of the surface, and refers to its plot only.

Fig. 4. Surface of R_p^* vs. measured values of the contents of the sausage components. Both the plots demonstrate total specific producer's risk R_p^* at the measured fat and protein contents constant and equal to the prior means. The measured moisture contents c_{3m} in Fig. 4a are on the interval of the three measurement standard uncertainties starting from the tolerance limit, while the interval of the measured salt contents c_{4m} is from the prior mean to the tolerance limit. In Fig. 4b the interval of the measured salt contents c_{4m} is of the three measurement standard uncertainties starting from the tolerance limit, whereas the interval of the measured moisture contents c_{3m} is from the prior mean to the tolerance limit. The color bars are as in Fig. 3.

Table 1. Results of ANOVA.

<i>i</i>	Comp.	Factory A		Factory B		$S_{i\text{between}}^2$ % ²	$S_{i\text{within}}^2$ % ²	F_i
		m_{iA} , %	S_{iA}^2 , % ²	m_{iB} , %	S_{iB}^2 , % ²			
1	Fat	40.9	5.44	40.1	22.81	14.63	13.37	1.09
2	Protein	25.0	1.77	24.1	1.79	17.50	1.78	9.84
3	Moisture	29.0	19.47	30.5	13.76	48.00	16.86	2.85
4	Salt	4.08	0.11	4.04	0.19	0.03	0.14	0.22

Table 2. Characteristics of the distributions.

<i>i</i>	Comp.	c_{min} , %	c_{max} , %	m_i , %	s_i , %	D_i	P_i
1	Fat	32.7	49.6	40.5	3.66	0.099	0.40
2	Protein	20.9	26.9	24.6	1.40	0.132	0.11
3	Moisture	22.6	44.5	29.7	4.15	0.089	0.52
4	Salt	2.90	5.10	4.07	0.38	0.105	0.32

Table 3. Matrix of empirical correlation coefficients r_{ij} .

<i>i / j</i>	Comp.	Fat	Protein	Moisture	Salt
		1	2	3	4
1	Fat	1.000	-0.163	-0.318	-0.217
2	Protein	-0.163	1.000	-0.235	0.301
3	Moisture	-0.318	-0.235	1.000	-0.111
4	Salt	-0.217	0.301	-0.111	1.000

Table 4. Covariance cov_{ij} matrix of the raw data.

i/j	Comp.	Fat	Protein	Moisture	Salt
		1	2	3	4
1	Fat	13.39	-0.84	-4.83	-0.30
2	Protein	-0.84	1.97	-1.37	0.16
3	Moisture	-4.83	-1.37	17.24	-0.17
4	Salt	-0.30	0.16	-0.17	0.14

Table 5. Matrix of correlation coefficients r_{ijco} after the closure operation.

i/j	Comp.	Fat	Protein	Moisture	Salt
		1	2	3	4
1	Fat	1.000	-0.142	-0.823	-0.165
2	Protein	-0.142	1.000	-0.436	0.511
3	Moisture	-0.823	-0.436	1.000	-0.230
4	Salt	-0.165	0.511	-0.230	1.000

Table 6. Covariance cov_{ijco} matrix of the prior pdf after the closure operation.

i/j	Comp.	Fat	Protein	Moisture	Salt
		1	2	3	4
1	Fat	10.63	-0.83	-9.56	-0.24
2	Protein	-0.83	3.22	-2.79	0.40
3	Moisture	-9.56	-2.79	12.71	-0.36
4	Salt	-0.24	0.40	-0.36	0.19

Table 7. Covariance cov_{ijlf} matrix of the multivariate likelihood function.

i/j	Comp.	Fat	Protein	Moisture	Salt
		1	2	3	4
1	Fat	4.11	-0.32	-1.15	-0.07
2	Protein	-0.32	0.96	-0.41	0.05
3	Moisture	-1.15	-0.41	3.18	-0.03
4	Salt	-0.07	0.05	-0.03	0.03

Declaration of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.



[Click here to access/download](#)

MethodsX
RawData.txt



Click here to access/download

MethodsX
R_code.R

Possible reviewers

1. Prof. Tony R.D.L. Dadamos, Fundação Educacional de Fernandópolis, Brazil, tonyrlrl@yahoo.com.br
2. Prof. Alexander Yu. Prosekov, The Kemerovo State University, Russia, aprosekov@rambler.ru
3. Dr. Angelique Botha, National Metrology Institute of South Africa (NMISA), South Africa, ABotha@nmisa.org
4. Dr. Alexandre Allard, The Laboratoire national de métrologie et d'essais (LNE) – The National Laboratory of Metrology and Testing, France, alexandre.allard@lne.fr