Expert system of food sensory evaluation for mobile and tablet

M A Nikitina[1], **Y A Ivashkin**[2]

[1]V.M. Gorbatov Federal Research Center for Food Systems of Russian Academy of Sciences, Talalikhina str., 26, Moscow, Russia, 109316
[2]Moscow Technical University Communication and Informatics, Aviamotornaya str., 8a, Moscow, Russia, 111024

e-mail: nikitinama@yandex.ru

Abstract. One of the main directions of statistics in sensory evaluation is an assessment of the dependence between experimental variables and measured characteristics. Statistical criteria are used to assess a degree of interaction between variables, a level of experimental effects, and allow accepting or rejecting hypothesis proposed. In sensory evaluation, people act as measurement instruments, and a variation associated with the human factor arises. This proves that the use of statistical methods is necessary. This article represents a network computer system for collection and evaluation of food sensory indicators based on the methods of rank correlation and multifactorial analysis of variance in real time. The article describes information technology of expert sensory evaluation of food quality by individual panelists and sensory panels regarding the indicators that are not measured by technical means of control, based on client-server network architecture. The software implementation of system for collecting and statistical processing of sensory data based on the principles of multifactorial analysis of variance in real-time mode makes it possible to evaluate the influence of the human factor on objectiveness and reliability of sensory evaluation results, as well as to visualize the data of expert scores by various expert panels.

1. Introduction

Sensory analysis and evaluation of food quality is the basis for commodity examination of food products and prediction of consumer demand. The usage of modern methods of sensory analysis requires from panelists not only specialized knowledge of methodology and application of, procedures for generating lexical dictionaries or scaling, but also the provision of consolidated, consistent scores that confirm the objectiveness of the results [1].

Regardless of the experience and degree of training of panelists, individual differences always arise in their scores of food quality, associated with sensory sensitivity, knowledge, and the subjectivity of the scales for sensory indicator measurement [2, 3]. Therefore, the adequate interpretation and objectification of individual panelist scores requires the development of IT technologies for factor analysis of food sensory evaluation results based on mathematical statistics methods [4-6] and computer software for processing and visualizing data from sensory evaluation of a product [7-8].

An objective assessment of food quality, taking into account a variety of parameters, alternatives and criteria, may be implement by using intelligent computer technologies for processing and formalizing knowledge with the adoption of optimal decisions based on statistical methods of multivariate data analysis with an objective assessment of adequacy and confirmation of hypotheses.

The probabilistic spread of characteristics and properties of initial biological raw material, as well as subjectivity of individual panelist opinions determine the complexity of this problem. The samples for testing cannot be identical in nutritional and biological value (protein, fat, connective tissue, etc.). If the expert does not have enough knowledge and competencies, then even the most sophisticated software will not conduct a qualitative analysis of the data [9].

Therefore, the adequacy of the scores in each case essentially depends on the influence of the human factor with the individual psycho-physiological capabilities of sensory perception and needs correction and checking the objectivity of expert scores, considering the coherence of certain panel opinions and experience [10-13].

Information technology in a client-server network architecture is proposed based on the methods of ranking correlation and multivariate analysis of variance, for sensory evaluation of food quality by individual panelists and sensory panels regarding the indicators that are not measured by any instruments [14].

2. Rank correlation in sensory evaluation of food quality

Formal characterization of final product sensory evaluation obtained as a result of expert survey is achieved by rank correlation [8,9,11,15,16], according to which a group of quantitatively non-measurable factors is ranked by each expert independently of each other in order of decreasing or increasing their influence on the assessment of product quality. The ranking results are recorded in a rank matrix x_{ij} , $i = \overline{1, m}$, $j = \overline{1, n}$, specifying the place of j-th parameter among n other parameters by i-th expert.

Since opinions of panelists not always match each other, total ranks are determined to obtain an objective assessment.

$$R_j = \sum_{i=1}^{m} x_{i,j}; \quad j = \overline{1, n}, \tag{1}$$

and the coefficient of concordance W is calculated, which characterizes the objective relationship between the scores of m independent panelists using the equation:

$$W = \frac{12 \cdot S(d^2)}{m^2 \cdot (n^3 - n)}, \tag{2}$$

$$S(d^2) = \sum_{j=1}^{n} \left[\sum_{i=1}^{m} x_{i,j} - \frac{1}{2} m(n+1) \right]^2, \tag{3}$$

ranging from 0 in the absence of relationship between rankings by experts, to 1 with their full consent in ranking the impact on the product quality Q.

After assessing the significance showing that the indicator of consistency W of panelist opinions is not accidental, weighting factor g_j is assigned to each sensory indicator:

$$g_j = \frac{M}{\sum_{i=1}^{m} x_{ij}}; \quad j = \overline{1, n}, \tag{4}$$

where M – scale factor.

Weighting factors g_j reflect the practical experience of qualified experts and characterize comparative impact of j-th factor on total quality assessment in regression:

$$Q = Q_0 + \sum_{j=1}^{n} g_j x_j, \tag{5}$$

where Q_0 – score of reference product, and allows to objectively identify the most significant factors of deviations from the specified quality.

The vector of weight coefficients is presents as a bar chart in accordance with the index number of the indicator or in descending order of the absolute values of the coefficients. If the distribution of weights is uniform or close to it, then the level of a priori knowledge is low and further accumulation and processing of statistical data is necessary. In turn, uneven distribution with an exponential decrease in the weights corresponds to a high degree of a priori knowledge about the product quality.

3. Multifactorial analysis of variance for sensory evaluation

The data with a multi-level structure is analyzed by multifactorial statistical procedures [17], which allow determining differences between two or more data sets for all dependent variables simultaneously. This helps to reduce the level of overall mistake of the first kind and to assess the degree of relationship between dependent variables, and makes it possible to establish combinations of sensory variables, which allow distinguishing samples in the case of non-manifestation of differences in each variable separately[18-28].

The general algorithm for processing the results of sensory evaluation includes the determination of the sample size from the general population and formulation of a null (H_0) and alternative (H_1) hypotheses; choice of significance level ($\alpha = 0.01$; or $\alpha = 0.05$; or $\alpha = 0.1$) and conducting an assessment; data collection and calculation of total statistical criteria; acceptance or rejection of the null (H_0) hypothesis and result interpretation.

In the case of two-factor analysis of variance, pooling by two factors is used, and in addition to the experimental error, the variance of scores due to individual differences between panelists in the panel is taken into account [29]. In this case, the order of samples A, B and C should be individual for each panelist, and their combinations ABC, ACB, BAC, BCA, CAB, CBA are distributed in equal proportions to ensure complete randomization. The total sample variance for all experiments is equal to sum of the intergroup and intragroup variances. The value of Fisher test is calculated not only as the ratio of the intergroup and intragroup variances of panelist scores, but also as the ratio of the variance of scores between individual experts to the intragroup variance.

Analysis of variance [17, 18] is the basis for software development with a client-server architecture for collection and statistical processing of sensory data. Fisher's exact test goes through all possible options of contingency table with the same total frequencies in rows and columns, i.e. carries out all kinds of construction of null-models, which built on the assumption of no influence on the factor under study [30-32].

The general algorithm for processing the results of sensory evaluation (Figure 1) includes determining the sample size from the general population and formulating null (H_0) and alternative (H_1) hypotheses; selecting the level of significance ($\alpha = 0.01$; or $\alpha = 0.05$; or $\alpha = 0.1$) and conducting an assessment; data collection and calculation of total statistical criteria; accepting or rejecting the null (H_0) hypothesis and interpreting the results.

The influence of the factor is estimated by the Fisher-Snedecor test at the chosen level of significance, not only as the ratio of the intergroup variance of scores from the relationship of factors to the intragroup variance from the experimental error $F = \frac{MS^2_{x1x2}}{MS^2_E}$, but also as a relationships $F = \frac{MS^2_{x1}}{MS^2_E}$ and $F = \frac{MS^2_{x2}}{MS^2_E}$ of factor variances of expert scores for intragroup variance.

The calculation of the average estimates of the i - th level X2 at the j - th level X2 and j - th level x at the i - th level X1 group of experts - tasters

$$\bar{y}_i = \frac{1}{n_2 q}\sum_{j=1}^{n_2}\sum_{k=1}^{q} y_{ijk}; \quad i = 1, n_1$$

$$\bar{y}_j = \frac{1}{n_1 q}\sum_{i=1}^{n_1}\sum_{k=1}^{q} y_{ijk}; \quad j = 1, n_2$$

and total average

$$\bar{y} = \frac{1}{n_1 n_2 m}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{k=1}^{q} y_{ijk}$$

Calculation of sums SSX1, SSX2 squared deviations estimates yi X1 and X2 from the average level and the whole

$$SS_{x1} = \sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{k=1}^{q}(\bar{y}_i - \bar{y})^2 = n_2 q\sum_{i=1}^{n_1}(\bar{y}_i - \bar{y})^2$$

$$SS_{x2} = \sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{k=1}^{q}(\bar{y}_j - \bar{y})^2 = n_1 q\sum_{j=1}^{n_2}(\bar{y}_j - \bar{y})^2$$

Calculation of MSX1, MSX2 dispersions

$$MS_{x1} = \frac{n_2 q}{n_1 - 1}\sum_{i=1}^{n_1}(y_i - \bar{y})^2 = \frac{SS_{x1}}{n_1 - 1}$$

$$MS_{x2} = \frac{n_1 q}{n_2 - 1}\sum_{j=1}^{n_2}(y_j - \bar{y})^2 = \frac{SS_{x2}}{n_2 - 1}$$

Figure 1. The integrated block diagram of two-factor analysis of variance.

4. Software implementation of expert system for food sensory evaluation with a client-server network architecture

The client-server expert software consists of two subsystems: the server software and the client software (Figure 2). The client software is installed on the user's computer and transmits requests to the server subsystem to process data and requests from clients and to return them back to user's computer.

The functional structure of the system includes six modules that enter parameters and evaluate product descriptors; creating a data set for analysis; sensory profile; comparison with reference; help (for user and administrator).

The list of parameters (Figure 3) determined by the purpose of the sensory evaluation, includes:

- the number of samples and evaluated descriptors;
- type of scale (structured or unstructured); structured five- and nine-point scales are used, according to which each indicator has 5 or 9 degrees of quality, respectively. According to a five-point scale: 5 – excellent quality; 4 – good; 3 – satisfactory; 2 – unsatisfactory, but acceptable; 1 – unsatisfactory.

Nine-point scale recommended in the V.M. Gorbatov Federal Research Center for Food Systems expands the range of sensory scores with the introduction of quantitative characteristics: 9 – optimum quality; 8 – very good; 7 – good; 6 – above average; 5 – medium; 4 or 3 – acceptable,

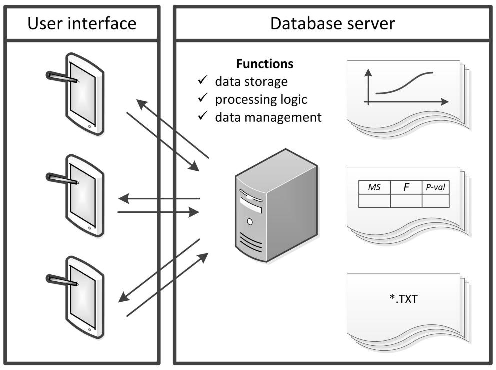

Figure 2. General scheme of the information system.

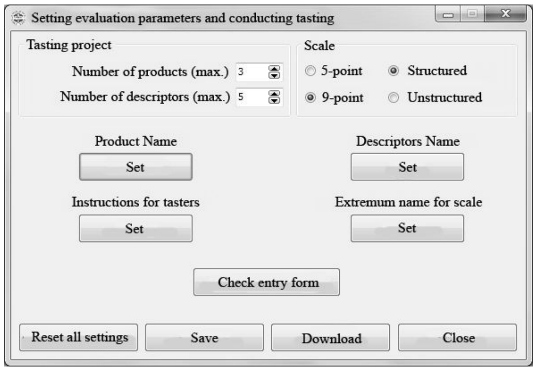

Figure 3. Start menu of the computer system.

but undesirable; 2 or 1 – unacceptable.

- name of the evaluated descriptors;
- folder for saving files with sensory evaluation results (text format *.txt);
- instruction for the sensory panel.

After setting the evaluation parameters, the panelist connects to the server software and enters his identification data (for example, full name) and further, using the intensity scale of the descriptors in the product samples, individually evaluates the intensity of the product descriptors recording the results from the beginning of the scale. After evaluating all the descriptors in the first product, the panelist proceeds with the next product or finishes the sensory evaluation.

Figure 4 represents a table with results of two-factor analysis of variance by the descriptor of "smoke odor".

As a null hypothesis (H_0), the system proposes: – tthe products do not affect the "smoke odor" tdescriptor, and the alternative hypothesis (H_1) – the products affect the "smoke odor" descriptor. To verify them, Fisher's exact test was used at a significance level of $\alpha = 0.05$.

From the data in Figure 5, the calculated value of *F-test* for x_1 factor (products) as $F \approx 19.85$, and the critical region is formed by the right-hand interval $(4.46; +\infty)$. Since F falls into the critical region, the null hypothesis (H_0) is rejected and the alternative (H_1) hhypothesis is accepted, i.e. x_1 factor (products) affects the "smoke odor".

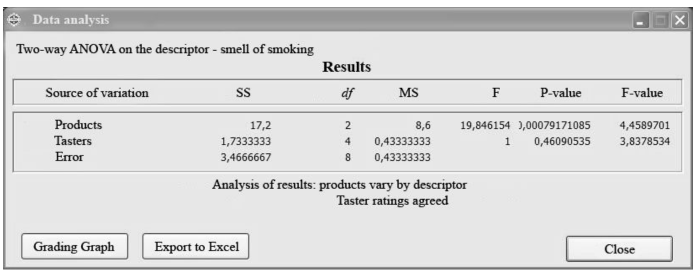

Figure 4. Result of the panelist's consistency check by the "Smoking odor": *SS* – variance; *df* – degree of freedom; *MS* – unbiased scores; *F* – calculated Fisher test; *P-value* – function of *F* distribution; *F critical value* – table value of Fisher test.

Similarly, assessment of the second factor takes place, i.e. "panelists". With a null (H_0) hypothesis, panelists do not affect the "smoke odor" descriptor and with alternative (H_1) hypothesis, panelists affect the "smoke odor" descriptor.

The values in Figure 5 show, that calculated F-test for x_2 factor (panelists) is $F = 1$, and the critical region is formed by the right-hand interval $(3.84; +\infty)$. Since F does not fall in the critical region, the null (H_0) hypothesis is accepted, i.e. influence of x_2 factor (panelists) on the "smoke odor" was not confirmed.

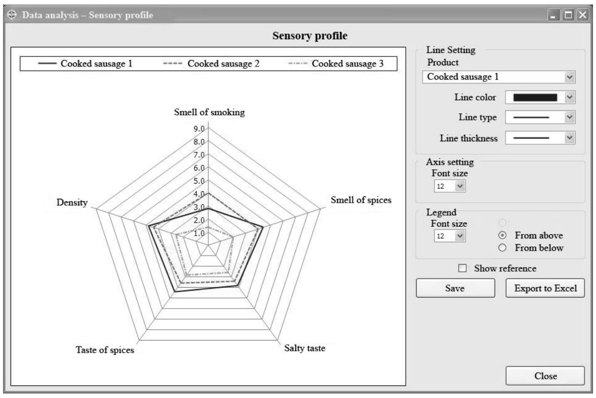

Figure 5. Sensory profile for three cooked sausage samples.

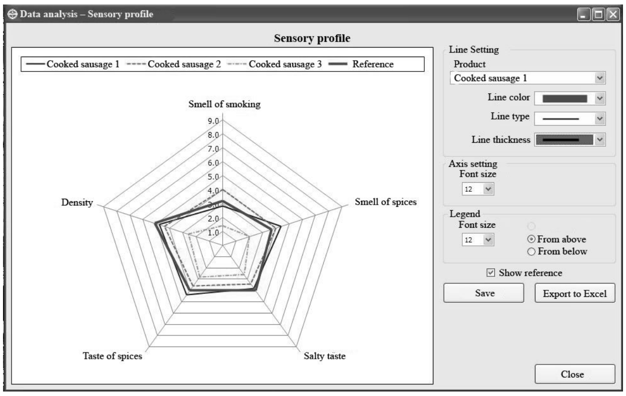

Figure 6. Sensory profile for the reference product and three cooked sausage samples.

Sample determination coefficient:

$$\overline{\rho}_{x_1^2} = \frac{SS_{x_1}}{SS_{x_1} + SS_{x_2} + SS_{x_\epsilon}} = \frac{17.2}{17.2 + 17.3 + 3.47} \approx 0.77, \tag{6}$$

shows that 77% of the total sample variation in the descriptor (smoke odor) is related to the influence of the product type on it.

In the *P-value* column P-value is determined, which corresponds to the calculated value of *F-test*.

In our example, *P-value* for x_1 factor (products) depends on the values of F; *df* and *MS* of this factor in the first row of the table, and has a value of 0.00079.

P-value for the x_2 factor (panelists) depends on the values of F; *df* and *MS* of this factor in the second row of the table, and is equal to 0.46.

According to the Fisher-Snedecor test, when *P-value* is less than 0.05 ($P < 0.05$), the data are not consistent. Based on the calculation, analysis and comparison, the system makes conclusion "Products differ in this descriptor; the scores of the panelists are consistent".

In the case of a consistent and reliable evaluation, the software allows to build a sensory profile (profilogram) of the product characteristic being evaluated (Figure 6) with a number of intensity score axes equal to the number of specific descriptors.

Figure 5 shows an example of the sensory profile for three samples of cooked sausage in the form of a polygon with vertices combining the obtained product characteristics.

Using similar procedures, the software allows to determine the position of a product among competitors based on a comparison of its profile with competitors' product profiles.

For comparison of the product profile with reference, reference product is preliminarily produced, which is the basis for comparing all the products involved in the evaluation. The characteristics of the reference sample determine the reference sensory profile, which is compared with the profile of a similar sample from another batch (Figure 6).

The computer software also allows to identify changes in the sensory characteristics of the product when replacing food ingredients, additives or spices in the formulation or using new types of packaging, etc.

5. Experimental testing of the software

The given example of the network expert system and its dialog interface along with the individual numerical scores and statistical evaluation by panelists provides the objectivized conclusion and recommendations concerning the product quality based on processing of the subjective data from expert panels of up to 20 panelists by 15 descriptors and 6 product types with the construction of profilograms with up to 15 descriptors and a possibility of data export to MS Excel. Thus, the accuracy and reliability of the objectivized scores presented in Figure 4 is determined by the criteria values for the specific case, as well as by the degree of agreement and competence of the opinions from the qualified panelists evaluating technically uncontrollable sensory properties of food products and their influence on evaluation.

6. Conclusion

Therefore, the computer software with the client-server architecture based on the multivariate analysis of variance realizes the information technology for support of decision making in sensory food evaluation contrary to the traditional expert systems and software packages. It performs real-time collection, accumulation and statistical processing of sensory data from individual panelists and geographically distributed panels and visual presentation of the objectified results in different graphic forms.

7. References

[1] Lawless H T and Heymann H 2010 *Sensory evaluation of food: principles and practices* (New York: Springer Science & Business Media) p 596

[2] Piggott J R and Hunter E A 1999 Review: Evaluation of assessor performance in sensory analysis Italian *Journal of Food Science* **4(11)** 289-303

[3] Tomic O, Forde C, Delahunty C and Naes T 2013 Performance indices in descriptive sensory analysis Acomplimentary screening tool for assessor and panel performance *Food Quality and Preference* **28(1)** 122-133

[4] Johnson R A and Wichern D W 2007 *Applied Multivariate Statistical Analysis* (New York: Prentice-Hall) p 800

[5] Kochan A and Grimm M 1985 Die sensorische Profilmethode *Lebensmittelindustrie* **32(6)** 255-269

[6] Surmacka-Szczesniak A 1998 Sensory texture proffiling-historical and scientific perspectives F*ood Technology* **52(8)** 14-16

[7] Labbe D, Rytz A and Hugi A 2004 Training is a critical step to obtain reliable product profiles in a real food industry context *Food quality and preference* **15(4)** 341-348

[8] Savela-Huovinen U, Muukkonen H and Toom A 2018 Sensory expert assessor's learning practices at workplace: Competencies and contexts in sensory evaluation *Journal of sensory studies* **33(2)** DOI: 10.1111/joss.12315

[9] Mitchell M N 2005 Strategically using General Purpose Statistics Packages: A Look at Stata, SAS and SPS *Technical Report Series. Statistical Consulting Group: UCLA Academic Technology Services* URL: http://www.ats.ucla.edu/stat/technical-reports/umber1_editedFeb_2_2007ucla_ATSstat_trl_1.1__207.pdf (08.09.2018)

[10] Running C A and Hayes J E 2016 Individual Differences in Multisensory Flavor Perception *Multisensory flavor perception: from fundamental neuroscience through to the marketplace* (*Woodhead Publishing Series in Food Science Technology and Nutrition)* **298** 185-210 DOI: 10.1016/B978-0-08-100350-3.00010-9.

[11] da Silva R D D N, Minim V P R, da Silva A N and Minim L A 2014 Number of judges necessary for descriptive sensory tests *Food quality and preference* **31** 21-27 DOI: 10.1016/j.foodqual.2013.07.010

[12] Turner R, Ledwith A and Kelly J 2012 Project management in small to medium-sized enterprises: Tailoring the practices to the size of company *Management Decision* **50(5)** 942-957 DOI: 10.1108/00251741211227627

[13] Yemelyanov S V, Burovoi I F and Levada F Yu 1998 Control of Indefinite Nonlinear dynamic systems. Induced internal feedback *Lecture Notes in Control and Information Sciences* **231** 196

[14] Brockhoff P B 2003 Statistical testing of individual differences in sensory profiling F*ood Quality and Preference* **14(5-6)** 425–434

[15] Naes T, Brockhoff P B andTomic O 2010 *Statistics for Sensory and Consumer Science* (John Wiley & Sons, Ltd) p 287 DOI: 10.1002/9780470669181
[16] Brand Z 2003 *Data analysis. Statistical and Computational Methods for Scientists and Engineers* (Moscow: Mir) p 686
[17] Stahle L and Wold S 1990 Multivariate analysis of variance (MANOVA) *Chemometrics and Intelligent Laboratory Systems* **9** 127-141
[18] Lee S M, Chung S J, Lee O H, Lee H S, Kim Y K and Kim K O 2008 Development of sample preparation, presentation procedure and sensory descriptive analysis of green tea *Journal of Sensory Studies* **23** 450-467 DOI: 10.1111/j.1745-459X.2008.00165.x
[19] Hoehl K, Schoenberger G U, Schwarz K and Busch-Stockfisch 2013 Is Perception of Sucrose and Caffeine Affected by Training or Experience? *Monitoring Training Effects in Female Subjects over a Half-Year Period* **28** 1–13
[20] Brockhoff P B and Sommer N A 2008 Accounting for scaling differences in sensory profile data *Proceedings of 10th European Symposium on Statistical Methods for the Food Industry* (Louvain-La-Neuve, Belgium) 283-290
[21] Smith A, Cullis B, Brockhoff P and Thompson R 2003 Multiplicative mixed models for the analysis of sensory evaluation data *Food Quality and Preference* **14(5-6)** 387-395
[22] Bi J 2003 Agreement and reliability assessments for performance of sensory descriptive panel *Journal of Sensory Studies* **18** 61-76
[23] Naima H A, Mehdi S and Morched C H 2013 Training and Control of a Tactile Sensory Panel *Journal of Applied Sciences* **13** 366-376
[24] Bi J and Kuesten C 2012 Intraclass Correlation Coefficient (ICC): A Framework for Monitoring and Assessing Performance of Trained Sensory Panels and Panelists *Journal of Sensory Studies* **27(5)** 352-364 DOI: 10.1111/j.1745-459X.2012.00399.x
[25] Grawford C M and Di Benedetto C A 2014 *New Products Management* (New York: McGraw-Hill Education) p 608
[26] Moskowitz H R, Porretta S and Silcher M 2005 *Concept research in food product design and development* (Ames: Blackwell Publishing) p 612
[27] Stone H, Bleibaum N R and Thomas A N 2012 *Sensory Evaluation Practices* (San Diego: Elsevier Academic Press) p 438
[28] Johnson R A and Wichern D W 2007 *Applied Multivariate Statistical Analysis* (New York: Prentice-Hall) p 800
[29] O'Mahony M 1986 *Sensory Evaluation of Food. Statistical Methods and Procedures* (New York: Marcel Dekker) p 488
[30] Patefield W M 1981 Algorithm AS 159. An efficient method of generating r x c tables with given row and column totals *Applied Statistics* **30(1)** 91-97
[31] Gotelli N J and Entsminger N J 2003 Swap algorithms in null model analysis *Ecology* **84** 532-535
[32] Strona G, Ulrich W and Gotelli N J 2018 Bi-dimensional null model analysis of presence-absence binary matrices *Ecology* **99(1)** 103-115 DOI: 10.1002/ecy.2043