Research Article Environmetrics

Received: 19 April 2014, Revised: 23 February 2015, Accepted: 5 March 2015, Published online in Wiley Online Library:

(wileyonlinelibrary.com) DOI: 10.1002/env.2341

Environmental odor perception: testing regional differences on heterogeneity with application to odor perceptions in the area of Este (Italy)

R. Arboretti Giancristofaro[a], S. Bonnini[b]*, L. Corain[c] and D. Vidotto[d]

One way to evaluate the hazard of environmental degradation of the quality of life in a place, because of bad odors, is to consider the heterogeneity of possible types and of possible sources of malodors. In the literature, the problem of comparing heterogeneities of types or of sources of odors in two geographical areas has not yet been dealt with the due attention. The main reason is that methodological proposals for tests for heterogeneity comparisons are very rare. We propose a permutation test based on the comparison of linear combinations of sampling indices of heterogeneity. The good power behavior, especially for small differences of the degrees of heterogeneity of the two compared areas, is proved through a Monte Carlo simulation study. The application of the test for heterogeneity comparisons on the data of the survey on odor perceptions in the region of Este (Italy) performed in 2010 shows the usefulness of the proposed methodology, which can be added, as a complementary analysis, to the classical established techniques for studying the environmental impact of odors like dispersion models, dynamic olfactometry, smell maps determination, and others.

Keywords: odor perception; odor heterogeneity; permutation test; categorical variables

1. INTRODUCTION

Environmental odors can affect the assessment of air quality (Cain, 1987), can be considered a warning agent of ambient pollution (Cain and Turk, 1985; Ames *et al*., 1993), and can influence moods and psychological health (Knasko, 1993; Schiffman *et al*., 1995). A study of odor perception as information processing task is performed by Dalton (1996). Malodor can be seen as serious environmental problem, especially when it is perceived as harmful and with possible negative effects on health. For this reason, Evans and Cohen (1987) consider malodor as environmental stressor. As a consequence, "problem-oriented" and "emotion-oriented" coping efforts occur (refer to Campbell, 1983).

Before considering the specific problem of comparing odor heterogeneities between two areas or regions, let us review the literature related to empirical studies on the environmental impact of odors. Typical problems in the literature specialized in environmental stress from odors are exposure estimation, measurement of odor annoyance, and analysis of the relation between exposure estimation and annoyance (Cavalini *et al*., 1991).

For assessing the odor impact, a very common and increasingly used approach, especially to predict areas of odor nuisance near animal production facilities, is based on atmospheric dispersion models (Sheridan *et al*., 2004; Sarkar *et al*., 2003a; Boeker *et al*., 2000; Drew *et al*., 2007). Materials released into the atmosphere are carried by the wind and diluted by turbulence. As a consequence, a plume of polluted air is produced. This plume has the shape of a cone with the apex towards the source and is usually represented by a mathematical Gaussian model (Carney and Dodd, 1989; Smith, 1995) because a Gaussian distribution of odor concentration is assumed. A limit of this approach is represented by the lack of model validation (Sheridan *et al*., 2004). Furthermore, some of the models require the use of many years of hourly data to obtain valid estimates of short-term hourly peaks or of longer time averages. Another limit is the complexity of specification and the use of these models, whose inputs are topography, meteorological data, odor emission rates, and source characteristics like building heights, chimney heights and locations, temperature, and others.

An alternative approach for evaluating the odor impact is to use field measurements by trained panelists (van Langenhove and van Broeck, 2001). In this case, a group of trained observers, called the *sniffing team*, determines the region of odor perception and contributes to the assessment of emission rate and to determine the maximum odor perception distance. Nicolas *et al*. (2006) propose a method for odor assessment based on the joint use of sniffing team and dispersion model. A similar approach, within release experiments, is applied by Bilsen

* *Correspondence to: S. Bonnini, Department of Economics and Management, University of Ferrara, Via Voltapaletto 11, 44121 Ferrara, Italy.* *stefano.bonnini@unife.it*

a *Department of Territory and Agroforestal Systems, University of Padua, Padua, Italy*

b *Department of Economics and Management, University of Ferrara, Ferrara, Italy*

c *Department of Engineering and Management, University of Padua, Padua, Italy*

d *Department of Statistical Science, University of Padua, Padua, Italy*

and De Fré (2009). A possible drawback of the method based on a sniffing team is the fact that it is time-consuming, expensive, and difficult to apply in the presence of variable local weather conditions. The engagement of people, especially within the local community, for the measurement of odor impact, is typical of studies focused on the odor perception (Sarkar *et al*., 2003a).

A complementary method, based on the chemical nature of the phenomenon, is the gas chromatography–mass spectrometry (Kleeberg *et al*., 2005). This approach is commonly used in solid waste disposal landfill areas to estimate chemical composition of the emission and the concentrations of the main compounds. For establishing odor concentration, dynamic olfactometry, performed in specialized laboratories by considering the chemical nature of odors and using a panel of sniffers, is also commonly used (Romain *et al*., 2008). The result of an olfactometric measurement of an air sample is the odor concentration, expressed in odorimetric units per cubic meter (OU m^{-3}). A concentration of 1 OU m^{-3} corresponds to the perception threshold, that is to the situation where 50% of persons who sniff the sample perceive a smell and the remaining 50% of the persons do not perceive any smell. Van Langenhove and De Bruyn (2001) develop a procedure based on the joint use of olfactometry and sniffing team to determine odor emissions of intensive animal farming houses. The poor versatility of the olfactometric method represents its main defect and the main reason of its limited application, especially with low concentrations (Littarru, 2007).

An alternative way for the estimation of disturbance from odors is the research of odorous molecules in analytic way, that is with chemical methods. Instead of (or together to) field measurements by trained panelists, sometimes, electronic measurements are used in the environment under study. To this purpose, sensorial analyzers, called "electronic noses," are used. These tools include a chemical sensor matrix and an informatic system (structured as a neural network) for emulating the human olfactive system. The electronic nose modifies the gaseous mixture into electronic signals and correlates measurements and properties to classify the odors, through a process of odor recognition by comparing the results of new measurements with previously stored data. The main limits of this method are the impossibility of detecting odors whose olfactive threshold is very low and the determination of possible synergic, masking, or exalting effects of some mixing of substances. Littarru (2007) proposes a method for the study of odor nuisance based on the combined use of dynamic olfactometry and electronic nose.

Several works in the specialized literature are dedicated to the study of the relation between dispersion (computed exposure) and perception (Sarkar *et al*., 2003b; Sarkar and Hobbs, 2002). Some of these studies focus on the determination of the minimum separation distance between possible sources of malodors (agricultural enterprises, solid waste landfills, plants for treatment of waste, etc.) and residential areas (Nicolas *et al*., 2008; Guo *et al*., 2004; Schauberger et al., 2001). These works are very important for guidelines about land use planning.

Other research areas, related to odor perceptions, study perceptual relations between odors and use multidimensional scaling methods for determining smell maps (Carrie *et al*., 1999; Dawes *et al*., 2004).

One way to evaluate the hazard of degradation of the quality of life in a place, because of bad odors, is to consider the heterogeneity of possible types and of possible sources of malodors. This is especially crucial in small areas with several possible sources of malodors like farms, cultivated lands, industrial plants, and waste treatment plants. that can cause different possible types of odors. In this case, rather than dispersion models, chemical methods, olfactometry, or electronic noses, field measurements by trained sniffers represent the most suitable method for measuring the disturbance from odors. The main reason of using sniffing teams is the impossibility of other methods of detecting odors at very low concentrations. Another reason is related to the difficulty of distinguishing different types of odors, also considering that bad odors are related to chemical composites. Finally, the use of trained individuals is valuable, given that the goal of the study is the analysis of the quality of life. The typical study to which we refer is, for example, that described in Blumberg and Sasson (2001), regarding the recognition of bad odors through resident's reports.

In the present paper, the focus is on the inferential problem of testing the equality of the heterogeneities of types of perceived odors and of possible sources of odors in two regions. In particular, the goal is to test whether the heterogeneity of one region is greater than the heterogeneity of the other. To this aim, we propose a variant of the nonparametric test for heterogeneity comparisons in the presence of categorical data, proposed by Arboretti Giancristofaro et al. (2009). Our proposal is based on the joint use of different indices for computing the test statistic as linear combination of such indices of heterogeneity. In Section 2, the testing procedure is described. In Section 3, the results of Monte Carlo simulations for studying the power behavior of the test are shown. Section 4 is dedicated to the application of the method to a real problem related to a survey on the perceived odors of an area in the north of Italy. Section 5 contains some final remarks.

2. PERMUTATION TEST ON HETEROGENEITY OF ODOR PERCEPTIONS

Let us consider a categorical response variable X, and let us suppose that it takes categories in $A = \{A_1,\ldots,A_K\}$ with unobserved probability distribution $Pr\{X = A_k\} = \pi_k$, $k = 1,\ldots,K$. A could be the set of types of odors, the set of possible sources of odors, or similar characteristics. An index for measuring the degree of heterogeneity must satisfy the following properties:

a. It takes its minimum value when the distribution is degenerate.
b. It assumes increasingly greater values when moving away from the degenerate towards the uniform distribution.
c. It takes its maximum value when the distribution is uniform.

Several indices satisfy the three properties and can be used as a measure of the degree of heterogeneity. One of the most commonly used is the Gini index (Gini, 1912):

$$\tau_G = \sum_{k=1}^{K} \pi_k(1 - \pi_k) \quad (1)$$

Another very common index, in particular in the theory of information, is the Shannon entropy (Shannon, 1948):

$$\tau_S = -\sum_{k=1}^{K} \pi_k \log(\pi_k) \quad (2)$$

where log(x) is the natural logarithm of x and 0 log (0) is assumed to be equal to zero.

A general method for measuring heterogeneity is proposed by Rényi (1996) with the class of indices depending by the parameter δ:

$$\tau_{R_\delta} = \frac{1}{1-\delta}\log\left(\sum_{k=1}^{K} {\pi_k}^{\delta}\right) \quad (3)$$

The special subcases $\tau_{R_3} = -\frac{1}{2}\log\left(\sum_{k=1}^{K} {\pi_k}^{3}\right)$ and $\tau_{R_\infty} = -\log\max(\pi_1, \ldots, \pi_K)$ are considered in this paper.

Given two regions or domains, D_1 and D_2, and the problem of comparing the heterogeneities of odor reports (obtained from the sniffers), e.g., the heterogeneities of the types of odors or the heterogeneities of the possible sources of odors, let us denote with $Het(D_1)$ and $Het(D_2)$, their corresponding heterogeneities, and with τ_j, a measure of $Het(D_j)$, $j = 1,2$. The null hypothesis of the problem is

$$H_0 : [Het(D_1) = Het(D_2)] \equiv [\tau_1 = \tau_2] \quad (4)$$

and the alternative is

$$H_1 : [Het(D_1) > Het(D_2)] \equiv [\tau_1 > \tau_2] \quad (5)$$

Let us denote the probability that A_k is observed in D_j with $Pr\{X(D_j) = A_k\} = \pi_{jk}$, $j = 1,2$, $k = 1,..,K$. If the order of $\{\pi_{j1}, \pi_{j2}, \ldots, \pi_{jK}\}$ were known, the probabilities could be arranged in nonincreasing order like in the Pareto diagram: $\pi_{j(1)} \geq \pi_{j(2)} \geq \ldots \geq \pi_{j(K)}$, with $j = 1,2$. Considering that heterogeneity is related to the concentration of probability, the null hypothesis of equality in heterogeneity could be written as the equality of the Pareto diagrams:

$$H_0 : \pi_{1(k)} = \pi_{2(k)}, k = 1, \ldots, K \quad (6)$$

Similarly, the alternative hypothesis could be defined by comparing the cumulative sums of the ordered probabilities as follows:

$$H_1 : \sum_{k=1}^{s} \pi_{1(k)} \leq \sum_{k=1}^{s} \pi_{2(k)}, s = 1, \ldots, K-1 \quad (7)$$

and the strict inequality holds for at least 1 $s \in \{1, \ldots, K-1\}$.

Trivially, we have that $\sum_{k=1}^{K} \pi_{1(k)} = \sum_{k=1}^{K} \pi_{2(k)} = 1$, and for this reason, in the alternative hypothesis, the case $s = K$ is not considered.

This problem presents similarity with the comparison of population diversities (refer to Pardo, 2006; Patil and Taille, 1982). Even if from the environmental point of view, they are substantially distinct problems, from the statistical point of view, they are similar. As a matter of fact, large number of species and abundance of those species (great diversity) correspond to high heterogeneity of the categorical variable representing the species.

The specification of the hypotheses follows the same rules of any other testing problem. Instead of comparing means, variances, distributions, etc., we compare heterogeneities; thus, we can represent the hypotheses in terms of comparisons between two indices of heterogeneities. The directional test (where the alternative hypothesis is $\tau_1 > \tau_2$ or $\tau_1 < \tau_2$), that is dominance in heterogeneity, is more difficult to be solved (refer to Arboretti Giancristofaro, Bonnini and Pesarin, 2009). Other possible alternative hypotheses can be considered for this test but the equality ("=") must be included in the null hypothesis (refer to Pesarin and Salmaso, 2010).

The nonincreasing order of the probabilities is used to evaluate the concentration of the probability, that is the heterogeneity. Such concentration can be represented by the cumulative ordered probability (COP) $\sum_{k=1}^{s} \pi_{j(k)} (s = 1, \ldots, K)$, that is by the Pareto diagram. In the presence of maximum concentration of probability (minimum heterogeneity), the distribution is degenerate because one category is observed with probability equal to 1 and all the others with probability equal to 0; that is, we have the maximum probability equal to 1 and all the other probabilities equal to 0; thus, the COP is equal to 1 for each $s = 1,\ldots,K$. In the presence of minimum concentration of probability (maximum heterogeneity), the distribution is uniform; that is, we have all the probabilities equal to K^{-1}, and the COP is $\sum_{k=1}^{s} \pi_{j(k)} = \frac{s}{K}$, and it is equal to 1 only when $s = K$. In general, in the intermediate situations, if the COP of a distribution is less than or equal to the COP of another distribution, then the heterogeneity of the former is greater than the heterogeneity of the latter. Hence, for comparing the heterogeneities of two distributions, we can compare the COP of the two distributions. In this sense, the problem is similar to the test for stochastic dominance: under the null hypothesis, the two COPs are equal, then exchangeability between the variables, transformed according to the Pareto diagram rule, holds. In other words, the equality in heterogeneity between $X(D_1)$ and $X(D_2)$ is equivalent to the equality in distribution of $Y(D_1)$ and $Y(D_2)$, where $Y(D_j)$ is the transformation of $X(D_j)$ according to the Pareto diagram rule. Formally,

$$Y(D_j) = r \text{ if and only if } X(D_j) = A_k \text{ and } r = rank(\pi_k) \text{with } r = 1, \ldots, K \quad (8)$$

Hence, exchangeability under H_0 holds only for the new transformed Y variable and not for the original X variable.

Let us consider two samples of data of size n_1 and n_2 from D_1 and D_2, respectively, and let $\{f_{jk}; j = 1, 2, k = 1, \ldots, K\}$ be the observed $2 \times K$ contingency table and $\{f_{j(k)}; j = 1, 2, k = 1, \ldots, K\}$ be the table of the ordered frequencies, where f_{jk} is the absolute frequency of A_k in the jth sample and $f_{j(k)}$ is the kth ordered frequency in the jth sample. In the original contingency table, the marginal frequency of the kth column (category A_k) is $f_{.k} = f_{1k} + f_{2k}$, while the marginal frequency of the jth row is the sample size n_j. In the table of the ordered frequencies, the marginal frequency of the kth column (rank k) is $f_{.(k)} = f_{1(k)} + f_{2(k)}$, while the marginal frequency of the jth row is the sample size n_j. A permutation of the dataset consists in reassigning some observations of sample 1 to sample 2 and vice versa. In case of permutations of the dataset transformed with the Pareto rule, the corresponding table (permuted table) has the same marginal frequencies of the observed table of ordered frequencies, because neither the absolute frequencies of values 1,…,K in the pooled dataset nor the sample sizes change with the permutations. In other words, if $\left\{f^*_{j(k)}; j = 1, 2; k = 1, \ldots, K\right\}$ denotes a permuted ordered table, we have $f^*_{\cdot(k)} = f^*_{1(k)} + f^*_{2(k)} = f_{\cdot(k)}$ for $k = 1, \ldots, K$ and $\sum_{k=1}^{K} f^*_{j(k)} = n_j$ for $j = 1, 2$.

According to the definition of H_1 in terms of cumulative ordered probabilities, heterogeneity comparisons could be performed by comparing the COPs. In this sense, the problem is similar to the test for stochastic dominance in the presence of ordered categorical variables, where the cumulative distribution functions of the variables transformed according to the Pareto diagram rule, $Y(D_1)$ and $Y(D_2)$, are compared. Testing procedures for stochastic dominance problems have been proposed by Hirotsu (1986), Lumley (1996), Loughin and Scherer (1998), Nettleton and Banerjee (2001), Han *et al*. (2004), Loughin (2004), Agresti and Klingenberg (2005). The restricted maximum likelihood ratio test is one of the most common solutions (Wang, 1996; Cohen *et al*., 2000; Silvapulle and Sen, 2005), and it is based on mixtures of chi-squared asymptotic distributions for the test statistic. The main limit of these solutions is that the weights of the mixture depend on the unknown population distribution. Nonparametric solutions are proposed by Pesarin (1994), Brunner and Munzel (2000), Pesarin (2001), Troendle (2002), and Pesarin and Salmaso (2006).

If the true ordering of the probabilities were known, we could apply a permutation testing procedure similar to that for stochastic ordering problems because under the null hypothesis, exchangeability would hold. The fact that the ordered probabilities are unknown implies the need of estimating them by means of the ordered frequencies:

$$\hat{\pi}_{j(k)} = \frac{f_{j(k)}}{n_j} \tag{9}$$

Hence, we use a data-driven ordering that may differ from the true one and that presents sampling variability. The main consequence is that under the null hypothesis, exchangeability is not exact but only approximated. According to the Glivenko–Cantelli theorem (Shorack and Wellner, 1986), for large sample sizes, data-driven and true ordering are equal with probability 1; thus, exchangeability holds only asymptotically.

A reasonable test statistic for the problem may be

$$T_\tau = \hat{\tau}_1 - \hat{\tau}_2 \tag{10}$$

where $\hat{\tau}_j = \tau\left(\hat{\pi}_{j(1)}, \ldots, , \hat{\pi}_{j(K)}\right)$ is a sampling index of heterogeneity computed for the jth sample. For example, the test statistic based on the index of Gini is

$$T_G = \sum_{k=1}^{K}\left[\left(\frac{f_{2(k)}}{n_2}\right)^2 - \left(\frac{f_{1(k)}}{n_1}\right)^2\right] \tag{11}$$

the test statistic computed with the index of Shannon is

$$T_S = \sum_{k=1}^{K}\left[\left(\frac{f_{2(k)}}{n_2}\right)\log\left(\frac{f_{2(k)}}{n_2}\right) - \left(\frac{f_{1(k)}}{n_1}\right)\log\left(\frac{f_{1(k)}}{n_1}\right)\right] \tag{12}$$

the test statistic related to the index of Rényi of order 3 is

$$T_{R_3} = \frac{1}{2}\left[\log\sum_{k=1}^{K}\left(\frac{f_{2(k)}}{n_2}\right)^3 - \log\sum_{k=1}^{K}\left(\frac{f_{1(k)}}{n_1}\right)^3\right] \tag{13}$$

finally, the test statistic corresponding to the index of Rényi of order ∞ is

$$T_{R_\infty} = \log\left(\frac{f_{2(1)}}{n_2}\right) - \log\left(\frac{f_{1(1)}}{n_1}\right) \tag{14}$$

where $f_{j(1)}$ is the maximum absolute frequency observed in the jth sample.

By denoting with $T_{\tau(0)}$ the observed value of T_τ, the procedure of the permutation test for the problem under study is the following:

1. perform B independent permutations and, for the bth permutation, compute the permuted ordered table $\left\{f^*_{j(k)}; j = 1, 2; k = 1, \ldots, K\right\}$ and the corresponding permutation value of the test statistic: $T^*_\tau(b)$;
2. compute the permutation p-value as proportion of permutation values of the test statistic greater than or equal to the observed value: $\lambda_\tau = \sum_{b=1}^{B} I\left(T^*_{\tau(b)} \geq T_{\tau(0)}\right)/B$, where $I(x)$ is the indicator function that takes value 1 when x is true and value 0 otherwise;
3. reject the null hypothesis of equality in heterogeneity in favor of the alternative hypothesis if $\lambda_\tau < \alpha$ and do not reject the null hypothesis otherwise.

We recall that we are not interested in comparing the distributions of $X(D_1)$ and $X(D_2)$ but only their heterogeneities; thus, we consider the difference of the sampling indices of heterogeneity as test statistic. The observed proportions (relative frequencies) are sampling estimates of the unknown probabilities and are used to compute the mentioned indices. In the computation of the test statistic, the order is not important because each of the considered indices of heterogeneity is order invariant, that is $\tau_j = \tau(\pi_{j1}, \ldots, \pi_{jK}) = \tau(\pi_{j(1)}, \ldots, \pi_{j(K)})$. The order is important only for the transformation of the original X variables into new Y variables, according to the Pareto diagram rule.

The same testing procedure may be applied by using other test statistics. In particular, to obtain a test statistic not strictly dependent on a specific index τ, we propose to consider a linear combination of the four considered statistics, with weights given by the inverse of the maximum value of the corresponding indices. This weights ensure that all the combined indices have the same scale, that is the interval [0,1];

thus, the sampling differences of their normalized values can be summed, and the degree of importance of each index is not affected by the specific scale, where the index is defined. Formally,

$$T_{LC} = w_G T_G + w_S T_S + w_{R_3} T_{R_3} + w_{R_\infty} T_{R_\infty} = \boldsymbol{w}'\boldsymbol{T} \tag{15}$$

where $\boldsymbol{w} = \left(w_G, w_S, w_{R_3}, w_{R_\infty}\right)'$ is the vector of weights and $\boldsymbol{T} = \left(T_G, T_S, T_{R_3}, T_{R_\infty}\right)'$ is the vector of test statistics. The weights are

$$w_G = \frac{K}{K-1} \tag{16}$$

and

$$w_S = w_{R_3} = w_{R_\infty} = \frac{1}{\log K} \tag{17}$$

In other words, the test statistic T_{LC} is given by the sum of the test statistics based on the differences of the normalized sampling indices, $T_{LC} = T_{\widetilde{G}} + T_{\widetilde{S}} + T_{\widetilde{R_3}} + T_{\widetilde{R\infty}}$, where $T_{\widetilde{G}} = \widetilde{\tau}_{G,1} - \widetilde{\tau}_{G,2}$ denotes the difference of the normalized sampling indices of Gini and $T_{\widetilde{S}}, T_{\widetilde{R_3}}, T_{\widetilde{R\infty}}$ have similar meanings for the indices of Shannon and Rènyi.

It is worth noting that there is another important reason for using a combined test. Heterogeneity of categorical variables is not uniquely defined. Within the general definition of heterogeneity related to the "concentration" of probability, which can be evaluated through the Pareto diagrams, we consider, in the test statistic, more than one index for measuring heterogeneity. Literature proposes several indices of heterogeneity, and each of them considers a different aspect of heterogeneity. As a matter of fact, the four considered that the indices are not equivalent, because in some situations, heterogeneity of a population can be greater than another according to one index and less than the other according to a different index. Thus, the indices that are used to measure heterogeneity put into evidence different aspects of it. This is the reason why instead of considering only one index as test statistic, we prefer to combine the information of four test statistics, that is to combine four tests. The linear combination is just one possible choice of combining the four tests, not yet considered in the specific literature, alternative to the combination of the significance level functions commonly used. It is simple and reasonable because it is equivalent to the sum of four indices normalized in the [0,1] interval. The underlying idea is not to choose as a test statistic, a combined measure of heterogeneity according to the Euclidean geometry, but to combine the "information" provided by different tests, each informative on one aspect of the phenomenon under study. According to the Nonparametric Combination (NPC) methodology (refer to Pesarin, 2001), the proposed procedure consists in determining the multivariate permutation distribution of the four tests and in combining the permutation values of the test statistics through the linear combination, which in any case is "admissible" in the sense that it does not exist any other test, which is "uniformly better than it" (Pesarin, 2001). This is also because of the fact that

1) all permutation tests are conditional on the observed dataset;
2) pooled observed dataset is always sufficient for any underlying distribution in H_0;
3) the dimensionality of the minimal sufficient statistic is n, because except for very simple problems (only if data distribution belongs to regular exponential family the minimal sufficient statistic is unidimensional), the minimal sufficient statistic is the whole n-dimensional dataset;
4) does not exist any unidimensional test statistic containing the whole information in the dataset;
5) four specialized statistics contain more information than only one.

The result of combination is an univariate statistic that sums up the information on the multivariate distribution of the original test statistics, without specifying or assuming any distribution function and without explicitly formalizing the dependence structure of the partial tests (as usual within the likelihood methods). Hence, using a terminology typical of the NPC theory, we apply the direct combination to a multiaspect test (refer to Pesarin, 2001).

3. SIMULATION STUDY

To compare the power behavior of the tests $T_G, T_S, T_{R_3}, T_{R\infty}$, and T_{LC}, a Monte Carlo simulation study has been performed. Data of population j have been randomly generated by the random variable $Y_j = 1 + int[KU^{\gamma_j}]$, where U is a uniform random variable in the interval [0,1], $int(x)$ is the integer part of the real number x, and γ_j is a parameter denoting the degree of heterogeneity of population j. Random variable Y_j is discrete, and it takes the first K integer numbers, each of them representing one of the K categories. When $\gamma_j = 1$, Y follows a uniform distribution in the set $\{1, 2, \ldots, K\}$ and the heterogeneity is maximum. The more γ_j is far from 1, the lower the heterogeneity. Without loss of generality, we consider γ_j values in the interval [0, 1]. Trivially, when $\gamma_j = 0$, the distribution would be degenerate and the heterogeneity would be minimum. Hence, the greater γ_j in the interval [0, 1], the higher the heterogeneity. Several different simulation settings have been taken into account, considering the K values of 5 and 10, different sample sizes (in the balanced and unbalanced cases), and different values for γ_1 and γ_2, under the null and alternative hypothesis. For each setting, one thousand datasets have been generated and one thousand permutations have been performed to compute the p-values. The adopted method is not the only way for simulating data. We choose this formula because the use is very simple and the results can be easily interpreted.

In Figure 1, the rejection rates of the five tests under the null hypothesis of equality in heterogeneity ($\gamma_1 = \gamma_2$), as function of the degrees of heterogeneity γ_1 and γ_2, are plotted. It is evident that the power decreases as the heterogeneities increase. For low degrees of heterogeneity

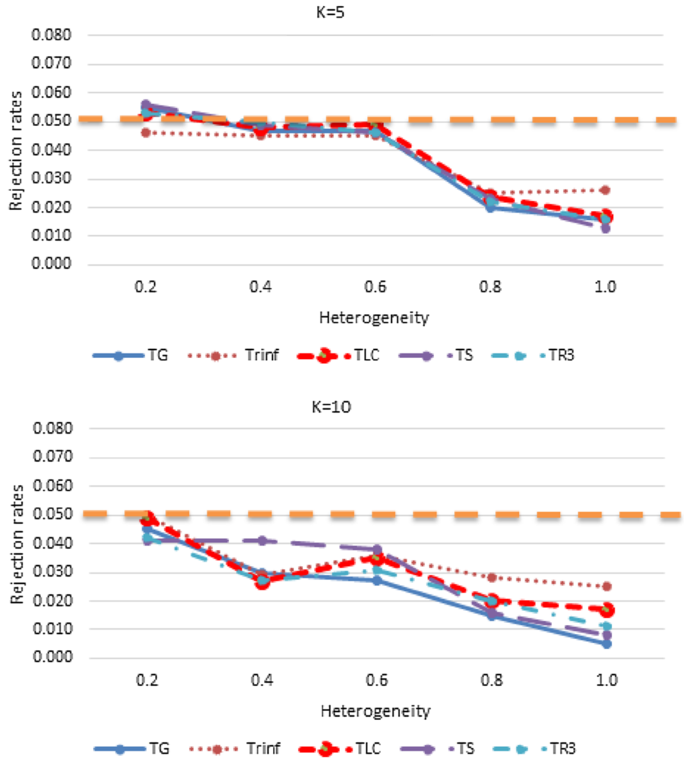

Figure 1. Rejection rates under the null hypothesis of equality in heterogeneity: $n_1 = n_2 = 60$, $\alpha = 0.05$, $B = 1000$

(specifically when $\gamma_1 = \gamma_2 = 0.2$), the estimated powers of some of the tests tend to be slightly greater than the significance level α but in general, we can say that all the tests all well approximated. When $K = 10$, the rejection rates are lower than in the case $K = 5$.

Figure 2 shows the rejection rates as function of the difference of the degrees of heterogeneity $\gamma_1 - \gamma_2$ (when $\gamma_1 = 1$). As expected, the

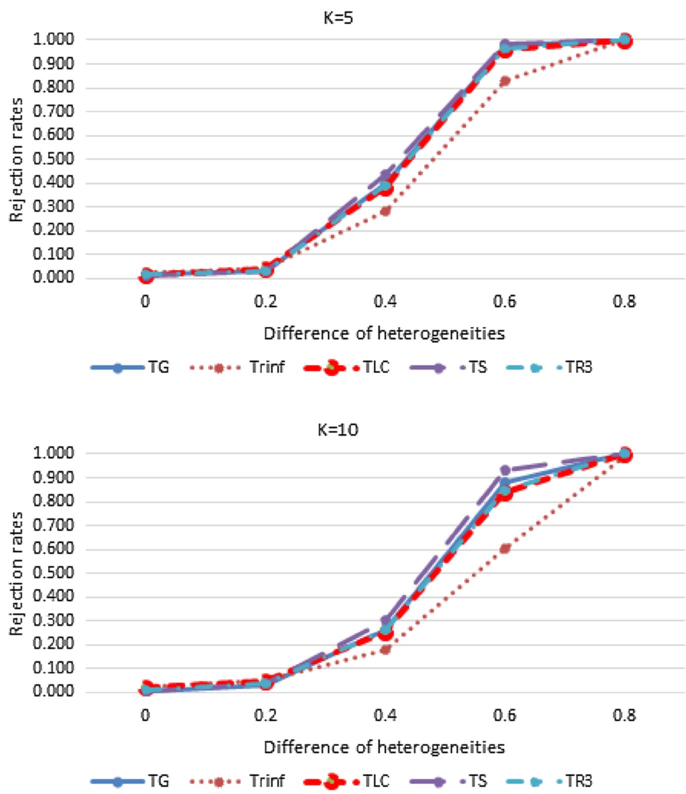

Figure 2. Rejection rates as function of the difference between the degrees of heterogeneity $\gamma_1 - \gamma_2 : \gamma_1 = 1$, $n1 = n2 = 60$, $\alpha = 0.05$, $B = 1000$

estimated powers are increasing function of the difference. When the difference is greater than 0.2, the growth rate of the power function is very high for all the tests. The performances of all the tests are very similar and very good. An exception to this rule is represented by the test $T_{R\,\infty}$, which seems to be less powerful than the others.

In Figure 3, the rejection rates are plotted as function of the sample sizes in the balanced case. According to the growth rates of the curves, we can say that the tests have a good power behavior in terms of consistency. Even in this case, the test based on the index of Rényi of order infinity shows the worst performance. Instead, the index of Shannon seems to be slightly preferable for large sample sizes.

Similar conclusions, about the comparative properties of the tests, can be drawn from the curves of the power as function of the difference between the sample sizes in the unbalanced case (Figure 4). By keeping fixed $n_1 + n_2 = 120$ and considering increasing values of the difference, starting from $n_1 = 20$ and $n_2 = 100$ ($n_1 - n_2 = -80$) up to $n_1 = 100$ and $n_2 = 20$ ($n_1 - n_2 = 80$), we can see that the power function reaches its maximum when the difference is zero, that is in the balanced case ($n_1 = n_2 = 60$).

Finally, in Figure 5, the rejection rates as function of the sample sizes in the balanced case are shown for the case $\gamma_1 = 1$ and $\gamma_2 = 0.8$, that is for small difference of the degrees of heterogeneity. In this peculiar setting, the test based on the linear combination of the indices seems to be the most powerful, together with the test based on the index of Rényi of order infinity. In the case $K = 10$, the power growth rate with the sample size of T_{LC} is the highest.

Hence, we proved the good power behavior of the test. Basically, the rejection rates under H_0 do not exceed the nominal alpha level. Of course, the test is not exact for finite sample sizes, but it is well approximated and asymptotically exact. We point out that most parametric methods are not exact. Many of them are based on asymptotic chi-squared distributions, and the approximation rate is not evaluated. Furthermore, the chi-squared test, commonly used for tests on contingency tables, is only asymptotically exact, and the convergence rate mainly depends on the unknown minimum probability. Usually, the approximation of the test is not taken into account by the researchers.

In the present work, we show the good approximation of the proposed test and the approximation is evaluated as a function of the heterogeneities, of the sample sizes, and of the number of categories.

4. APPLICATION EXAMPLE FOR THE ODOR PERCEPTIONS IN THE AREA OF ESTE (NORTH OF ITALY)

The application example considered in this paper is related to a statistical survey performed in the period of February 2010 to January 2011 in the area of Este, in the north-east of Italy. The goal of the survey was to determine the odor perceptions in that territory, characterized by the presence of several possible sources of malodors like farms, cultivated lands, and industrial and waste treatment plants. In particular, 81 trained sniffers (panelists), resident in the two municipalities of the area, Este and Ospedaletto Euganeo (refer to Figure 6), in the cited period, were asked to report perceived odors and to indicate, among other information, the type of perceived odor and the possible source.

This area is characterized by stable meteorological conditions, small extension, and presence of several possible sources of odors. To avoid possible confounding factors on the analysis, the period of the survey is divided into four seasonal periods (quarters) such that some

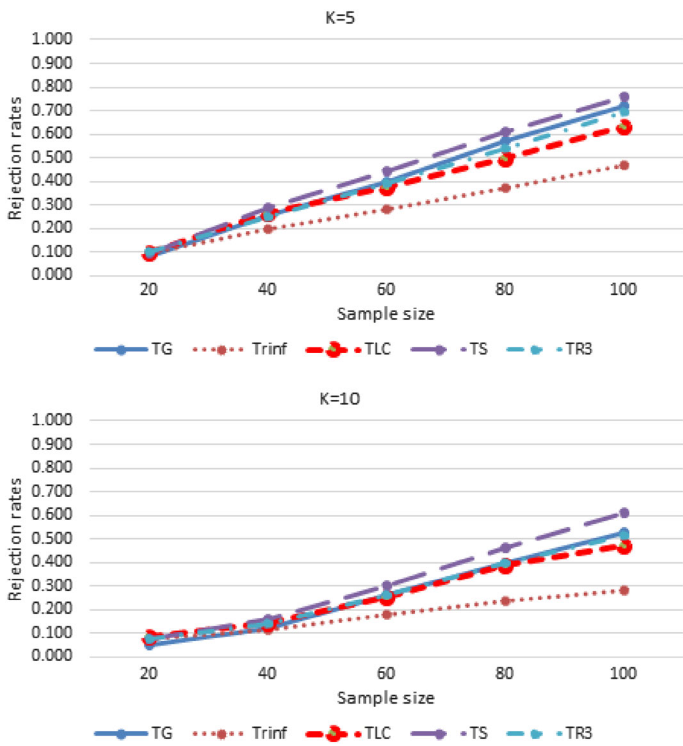

Figure 3. Rejection rates under the alternative hypothesis as function of the sample sizes (balanced case): $\gamma_1 = 1$ and $\gamma_2 = 0.6$, $\alpha = 0.05$, $B = 1000$

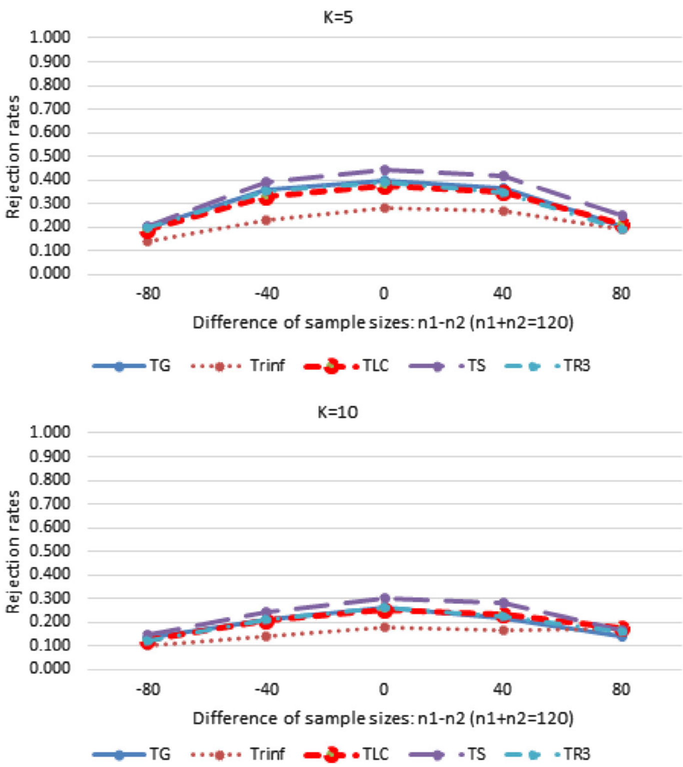

Figure 4. Rejection rates under the alternative hypothesis as function of the difference of sample sizes ($n_1 + n_2 = 120$): $\gamma_1 = 1$ and $\gamma_2 = 0.6$, $\alpha = 0.05$, $B = 1000$

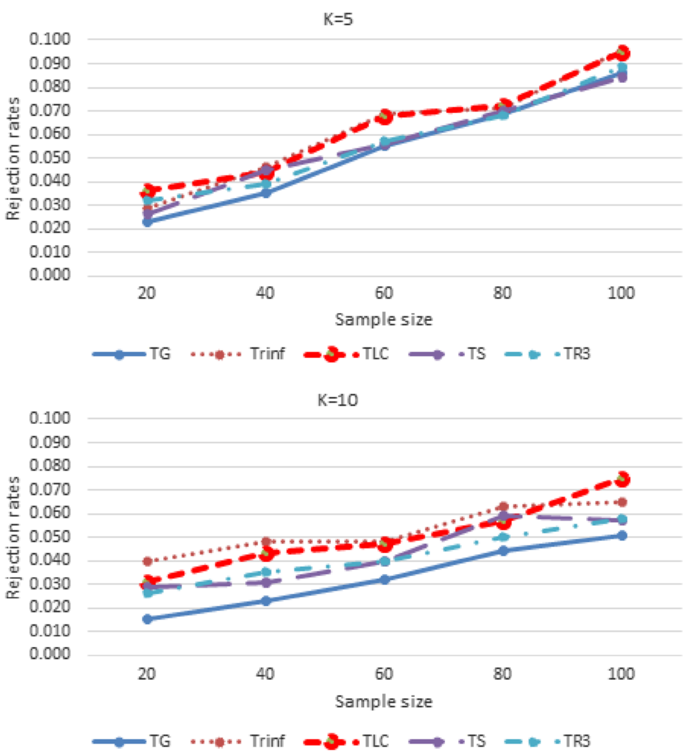

Figure 5. Rejection rates under the alternative hypothesis as function of the sample sizes (balanced case): $\gamma_1 = 1$ and $\gamma_2 = 0.8$, $\alpha = 0.05$, $B = 1000$

environmental conditions, like ambient temperature and types of activities in the territory, are homogeneous within each period. The quarters are defined in Table 1.

We are interested, for each quarter, in comparing the reports of odor perceptions of the two municipalities and test whether the heterogeneity of odor perceptions, in terms of types of odor and possible sources of odor, in Este are greater than in Ospedaletto Euganeo. Hence, the

Figure 6. Municipalities of Este and Ospedaletto Euganeo

Table 1. Subperiods in the survey on odor perceptions in the area of Este	
Period	Interval
Quarter 1	February 2010 to April 2010
Quarter 2	May 2010 to July 2010
Quarter 3	August 2010 to October 2010
Quarter 4	November 2010 to January 2011

hypotheses of the testing problem are

$$H_0 : Het(Este) = Het(Ospedaletto\ Euganeo) \tag{18}$$

against

$$H_1 : Het(Este) > Het(Ospedaletto\ Euganeo) \tag{19}$$

The contingency tables of the perceived odors in the four quarters are shown in Table 2 (types of odor) and Table 3 (possible sources of odor).

About the types of perceived odors, from the descriptive point of view, the distributions in Este and Ospedaletto Euganeo seem different. For example, the most frequent perceived odor in Este is "acrid-pungent" in the first and fourth quarters and "droppings" in the two central quarters, while in Ospedaletto Euganeo, only in the first quarter, the most perceived odor (acrid-pungent) coincides with that of Este, because in the other quarters, the mode is "putrid-rotten." About the possible sources of odors, the prevalent response is "waste treatment plant" in both the municipalities and in all the quarters, even if the frequency distributions do not seem similar. Because we are interested in comparing the heterogeneities of the distributions, let us consider the sampling indices of heterogeneity reported in Tables 4 and 5.

For the types of odors, indices of heterogeneities of Este tend to be greater than those of Ospedaletto Euganeo in the first and second quarters, while the opposite inequality is true in the third and fourth quarters. Hence, the greater heterogeneity of Este is not evident. For the possible sources of odors, in all the quarters, the indices related to Este take greater values than those related to Ospedaletto Euganeo.

By performing the permutation test for heterogeneity comparisons, it is possible to test whether Este has a greater heterogeneity than Ospedaletto Euganeo for both types and sources of odors, in all the four quarters. We highlight that in this study, like in the survey for monitoring bad odor dispersion in the northern Negev, described by Blumberg and Sasson (2001), the statistical unit is a single report of odor, characterized by a specific type of perceived odor and by a specific possible source. Hence, more than one report can be related to the same panelist, and reports coming from the same panelist could be assumed dependent. The panelist can be considered the block factor of the experiment or identifies a second level in a multilevel structure of data; thus, the idea of panelist effect is reasonable. However, we assume that there is no panelist effect for two main reasons.

First, for removing the possible panelist effect, we could design the experiment with a matching technique, where the same number of panelists is considered in the two areas and each panelist in one area provides the same number of reports (replicates) of a "similar" panelist in the other area, and analyze data like in a paired data problem. With this design, exchangeability under H_0 holds between the two corresponding observations in Este and Ospedaletto Euganeo for each couple of paired data. The possible difference could come only by the symbolic treatment effect represented by the area. In this survey, this approach cannot be applied because the reports are spontaneous initiatives of each sniffer whenever it perceives a smell. Hence, the number of reports can change with the sniffer and can be zero, one, or more than one. In this survey, there were other goals, and one of them was to determine the number of spontaneous reports in the two areas.

Table 2. Types of perceived odors: observed contingency tables

	Animal	Acrid-Pungent	Ammonia	Hay-Forage	Roasting	Droppings	Putrid-Rotten	Other
				Quarter 1				
Este	39	164	10	1	4	92	94	123
Ospedaletto E.	31	113	5	1	9	43	77	48
				Quarter 2				
Este	34	100	10	3	5	108	69	37
Ospedaletto E.	30	75	3	0	12	51	103	31
				Quarter 3				
Este	30	85	10	6	0	101	49	31
Ospedaletto E.	3	37	26	1	8	38	50	22
				Quarter 4				
Este	8	84	2	2	3	31	40	22
Ospedaletto E.	6	11	16	0	12	16	37	24

Table 3. Possible sources of odors: observed contingency tables

	Poultry farm	Cattle breeding	Pigs breeding	Manure	Dung	Waste treatment	Feed mill	Cement factory	Other	Trafic
					Quarter 1					
Este	45	12	6	4	8	119	6	15	81	30
Ospedaletto E.	26	1	18	7	4	195	14	7	20	2
					Quarter 2					
Este	53	18	2	3	10	70	1	30	53	9
Ospedaletto E.	15	6	25	1	8	170	10	0	23	1
					Quarter 3					
Este	37	16	1	1	8	67	0	9	53	9
Ospedaletto E.	5	2	15	0	13	143	14	0	7	3
					Quarter 4					
Este	15	3	1	0	0	57	0	21	12	21
Ospedaletto E.	6	1	4	0	3	74	12	1	1	0

Table 4. Sampling indices of heterogeneities for the types of odors

Index	Municipality	Quarter 1	Quarter 2	Quarter 3	Quarter 4
Gini	Este	0.780	0.783	0.776	0.724
	Ospedaletto E.	0.776	0.776	0.809	0.815
Shannon	Este	1.632	1.678	1.651	1.524
	Ospedaletto E.	1.660	1.644	1.760	1.808
Rényi (3)	Este	1.457	1.453	1.407	1.158
	Ospedaletto E.	1.401	1.406	1.598	1.587
Rényi (∞)	Este	1.167	1.221	1.128	0.827
	Ospedaletto E.	1.063	1.086	1.308	1.193

An alternative strategy could be assuming under H_0 exchangeability of the panelists (and not of the single reports) between the areas. This solution assumes that the possible differences observed in the reports are because of the area effect. But assuming the presence of a panelist effect is equivalent to thinking that some personal characteristics of the sniffers can affect the perceptions and that we are in presence of two or more confounding factors. In this case, the possible observed differences between the two areas may also be determined by other factors, in addition to the area of origin, and exchangeability under H_0 is not guaranteed.

Table 5. Sampling indices of heterogeneities for the possible sources of odors

Index	Municipality	Quarter 1	Quarter 2	Quarter 3	Quarter 4
Gini	Este	0.773	0.807	0.773	0.733
	Ospedaletto E.	0.540	0.546	0.482	0.454
Shannon	Este	1.762	1.824	1.690	1.544
	Ospedaletto E.	1.275	1.245	1.115	1.018
Rényi (3)	Este	1.347	1.565	1.387	1.177
	Ospedaletto E.	0.614	0.628	0.517	0.479
Rényi (∞)	Este	1.008	1.269	1.099	0.824
	Ospedaletto E.	0.411	0.421	0.345	0.321

Table 6. *P*-values of the tests for heterogeneity (H_1: $Het(Este) > Het(Ospedaletto\ E.)$)

	Test	Quarter 1	Quarter 2	Quarter 3	Quarter 4
Typesof odors	T_G	0.384	0.319	0.986	0.999
	$T_{R\infty}$	0.146	0.121	0.928	0.991
	T_{LC}	0.281	0.177	0.980	0.999
Possible sources of odors	T_G	0.000	0.000	0.000	0.000
	$T_{R\infty}$	0.000	0.000	0.000	0.000
	T_{LC}	0.000	0.000	0.000	0.000

Second, those involved in the survey are selected and well trained sniffers; hence, it is reasonable to assume that a possible sniffer effect is negligible and that reports from the same sniffer are approximately independent. By assuming that the specific geographic area is characterized by stable meteorological conditions and small extension and that the sniffers are well trained, possible differences in the perceptions of the two areas can approximately be attributed to the area effect. As the statistical unit is the single report, under the null hypothesis exchangeability of reports between the two areas at least approximately holds, thus permutation concern reports.

Let us consider the permutation test based on the index of Gini that based on the index of Rényi of order ∞ and the proposed index based on the linear combination and consider the significance level $\alpha = 0.01$. In Table 6, the *p*-values of the tests are shown.

All the tests related to the types of odors, in all the four quarters, lead to not reject the null hypothesis of equality in distribution because the *p*-values are greater than α. Hence, there is no empirical evidence in favor of the hypothesis of greater heterogeneity of the types of odors in Este. Despite this result, the possible sources of odors, according to the sniffers' perceptions, present a greater heterogeneity in Este: as a matter of fact, the *p*-values of all the tests are approximately zero; thus, they are less than α in all the quarters.

5. CONCLUSIONS

The aim of comparing the environmental impact of odors perceived in two distinct geographical areas has been addressed by means of a test that compares the heterogeneities of the reports of odors perceived by teams of trained sniffers operating in two areas.

The proposed testing procedure consists in the transformation of sample data according to the Pareto diagram rule and in performing a two sample permutation test on the transformed variable, similar to the permutation test for stochastic dominance alternatives. Instead of considering as test statistic the difference of suitable sampling indices of heterogeneity, we propose to use a linear combination of the sampling indices of Gini, Shannon, Rényi of order 3, and Rényi of order ∞, where the weights of the combination are the inverse of the maximum value of each index.

The simulation study proves the good power behavior of all the procedures. All of them are well approximated, consistent ,and with high power growth rate, with respect to the differences of the heterogeneities. The solution based on the linear combination of the indices shows similar performance to the other considered tests, but it is the most powerful when the difference of the degrees of heterogeneity of the two areas is small. Also, the test based on the Rènyi index of order ∞ satisfies this good property, but in several other settings, it is much less powerful than the other tests.

The application of the permutation tests for heterogeneity comparisons on the data of the survey on odor perceptions in the region of Este, in the north of Italy, shows that the odor perceptions in the municipality of Este are more heterogeneous than in the municipality of Ospedaletto Euganeo, from the point of view of the possible sources of odors but not in terms of types of perceived odors. This result is conformed in all the quarters of the year.

Acknowledgements

The authors wish to thank the municipalities of Este and Ospedaletto Euganeo for the data and the support to the survey. They wish also to acknowledge Professor Luigi Salmaso of the University of Padova for coordinating the research group on this activity.

REFERENCES

Agresti A, Klingenberg B. 2005. Multivariate tests comparing binomial probabilities, with application to safety studies for drugs. Applied Statistics **54**(4): 691–706.

Ames RG, Howd RA, Doherty L. 1993. Community exposure to a paraquat drift. Archives of Environmental Health **48**: 47–52.

Arboretti Giancristofaro R, Bonnini S, Pesarin F. 2009. A permutation approach for testing heterogeneity in two-sample problems. Statistics & Computing **19**: 209–216.

Bilsen I, De Fré R. 2009. Sniffing team measurements in Flanders: code of good practice. Water Practice & Technology **4**(2): 1–8.

Blumberg DG, Sasson A. 2001. Municipal hotlines and automated weather stations as a tool for monitoring bad odour dispersion: the northern Negev case. Journal of Environmental Management **63**: 103–111.

Boeker P, Wallenfang O, Koster F, Croce R, Diekmann B, Griebel M, Schulze LP. 2000. Agrartechnische Forschung **6**(4): 84–89.

Brunner E, Munzel U. 2000. The nonparametric Behrens–Fisher problem: asymptotic theory and small sample approximation. Biometrical Journal **42**: 17–25.

Cain WS. 1987. Indoor air as a source of annoyance. In Environmental Annoyance: Characterization, Measurement and Control, HS Koelega (ed.), Elsevier: Amsterdam.

Cain WS, Turk A. 1985. Smell of danger: an analysis of LP-gas odorization. American Industrial Hygiene Association Journal **46**: 115–126.

Campbell JM. 1983. Ambient stressors. Environment and Behavior **15**: 355–380.

Carney P, Dodd VA. 1989. The measurement of agricultural malodours. Journal of Agricultural Engineering Research **43**: 197–209.

Carrie S, Scannell JW, Dawes PJD. 1999. The smell map: is there a commonality of odour perception? Clinical Otolaryngology **24**: 184–189.

Cavalini PM, Koeter-Kemmerling LG, Pulles MPJ. 1991. Coping with odour annoyance and odour concentrations: three field studies. Journal of Environmental Psychology **11**: 123–142.

Cohen A, Kemperman JHB, Madigan D, Sarkrowitz HB. 2000. Effective directed tests for models with ordered categorical data. Australian and New Zealand Journal of Statistics **45**: 285–300.

Dalton P. 1996. Odor perception and beliefs about risk. Chemical Senses **21**(4): 447–458.

Dawes PJD, Dawes MT, Williams SM. 2004. The smell map: commonality of odour perception confirmed. Clinical Otolaryngology **29**: 648–654.

Drew GH, Smith R, Gerard V, Burge C, Lowe M, Kinnersley R, Sneath R, Longhurst PJ. 2007. Appropriateness of selecting different averaging times for modelling chronic and acute exposure to environmental odours. Atmospheric Environment **41**: 2870–2880.

Evans GW, Cohen S. 1987. Environmental stress. In Handbook of Environmental Psychology, D Stokols, I Altman (eds.), Wiley: New York; 571–610.

Gini C. 1912. Variability and Mutability. In Studi economico-giuridici della Facoltà di Giurisprudenza. University of Cagliari: Cagliari, Italy.

Guo H, Jacobson LD, Schmidt DR, Nicolai EE, Janni KA. 2004. Comparison of five models for setback distance determination from livestock sites. Canadian Biosystems Engineering **46**: 617–625.

Han KE, Catalano PJ, Senchauduri P, Mehta C. 2004. Exact analysis of dose-response for multiple correlated binary outcomes. Biometrics **4**(60): 216–224.

Hirotsu C. 1986. Cumulative chi-squared statistic as a tool for testing goodness-of-fit. Biometrika **73**: 165–173.

Kleeberg KK, Liu Y, Jans M, Schlegelmilch M, Streese J, Stegmann R. 2005. Development of a simple and sensitive method for the characterization of odorous waste gas emissions by means of solid-phase microextraction (SPME) and GC-MS/olfactometry. Waste Management **25**: 872–879.

Knasko SC. 1993. Performance, mood, and health during exposure to intermittent odors. Archives of Environmental Health **48**: 305–308.

Littarru P. 2007. Environmental odours assessment from waste treatment plants: dynamic olfactometry in combination with sensorial analyzers "electronic noses". Waste Management **27**: 302–309.

Loughin TM. 2004. A systematic comparison of methods for combining p-values from independent tests. Computational Statistics and Data Analysis **47**: 467–485.

Loughin TM, Scherer PN. 1998. Testing for association in contingency tables with multiple column responses. Biometrics **54**: 630–637.

Lumley T. 1996. Generalized estimating equations for ordinal data: a note on working correlation structures. Biometrics **52**: 354–361.

Nettleton D, Banerjee T. 2001. Testing the equality of distributions of random vectors with categorical components. Computational Statistics and Data Analysis **37**: 195–208.

Nicolas J, Craffe F, Romain AC. 2006. Estimation of odour emission rate from landfill areas using the sniffing team method. Waste Management **26**: 1259–1269.

Nicolas J, Delva J, Cobut P, Romain AC. 2008. Development and validating procedure of a formula to calculate a minimum separation distance from piggeries and poultry facilities to sensitive receptors. Atmospheric Environment **42**: 7087–7095.

Pardo L. 2006. Statistical Inference Based on Divergence Measures. Chapman & Hall/CRC: Boca Raton.

Patil GP, Taille C. 1982. Diversity as a concept and its measurement. Journal of the American Statistical Association **77**: 548–567.

Pesarin F. 1994. Goodness-of-fit testing for ordered discrete distributions by resampling techniques. Metron **52**: 57–71.

Pesarin F. 2001. Multivariate Permutation Test with Applications to Biostatistics. Wiley: Chichester.

Pesarin F, Salmaso L. 2006. Permutation tests for univariate and multivariate ordered categorical data. Austrian Journal of Statistics **35**: 315–324.

Pesarin F, Salmaso L. 2010. Permutation Tests for Complex Data: Theory, Applications and Software. Wiley: Chichester.

Rényi A. 1996. Calculus des Probabilitès. Dunod: Paris.

Romain AC, Delva J, Nicolas J. 2008. Complementary approaches to measure environmental odours emitted by landfill areas. Sensors and Actuators B **131**: 18–23.

Sarkar U, Hobbs SE. 2002. Odour from municipal solid waste (MSW) landfills: a study on the analysis of perception. Environment International **27**(8): 655–662.

Sarkar U, Hobbs SE, Longhurst P. 2003a. Dispersion of odour: a case study with a municipal solid waste landfill site in North London, United Kingdom. Journal of Environmental Management **68**: 153–160.

Sarkar U, Longhurst PJ, Hobbs SE. 2003b. Community modelling: a tool for correlating estimates of exposure with perception of odour from municipal solid waste (MSW) landfills. Journal of Environmental Management **68**: 133–140.

Schauberger G, Piringer M, Petz E. 2001. Separation distance to avoid odour nuisance due to livestock calculated by the Austrian odour dispersion model (AOMS). Agriculture, Ecosystems & Environment **87**: 13–28.

Schiffman SS, Miller EAS, Suggs MS, Graham BG. 1995. The effect of environmental odors emanating from commercial swine operations on the mood of nearby residents. Brain Research Bulletin **37**: 369–375.

Shannon CE. 1948. A mathematical theory of communication. Bell System Technology Journal **27**: 379–423.

Sheridan BA, Hayes ET, Curran TP, Dodd VA. 2004. A dispersion modelling approach to determining the odor impact of intensive pig production units in Ireland. Bioresource Technology **91**: 145–152.

Shorack GR, Wellner JA. 1986. Empirical Processes with Applications to Statistics. Wiley: New York.
Silvapulle MJ, Sen PK. 2005. Constrained Statistical Inference, Inequality, Order, and Shape Restrictions. Wiley: New York.
Smith RJ. 1995. A Gaussian model for estimating odor emissions from area sources. Mathematical Computing Modelling **21**(9): 23–29.
Troendle JF. 2002. A likelihood ratio test for the nonparametric Behrens–Fisher problem. Biometrical Journal **44**(7): 813–824.
Van Langenhove H, De Bruyn G. 2001. Development of a procedure to determine odour emissions from animal farming for regulatory purposes in Flanders. Water Science and Technology **44**(9): 205–210.
Van Langenhove H, Van Broeck G. 2001. Applicability of sniffing team observations: experience of field measurements. Water Science and Technology **44**(9): 65–70.
Wang Y. 1996. A likelihood ratio test against stochastic ordering in several populations. Journal of the Americal Statistical Association **91**: 1676–1683.

