Use of multidimensional testing to evaluate the impacts of treated wastewater discharge on river water quality – Hotelling test case

D Majerek¹, G Łagód² B Szeląg³ and F Sabba⁴
¹ Lublin University of Technology, Faculty of Fundamentals of Technology, Nadbystrzycka 38, 20-618 Lublin, Poland
² Lublin University of Technology, Faculty of Environmental Engineering, Nadbystrzycka 40B, 20-618 Lublin, Poland
³ Kielce University of Technology, Faculty of Environmental, Geomatic and Energy Engineering, Tysiąclecia Państwa Polskiego 7, 25-314 Kielce, Poland
⁴ Northwestern University, Department of Civil and Environmental Engineering, Evanston, Illinois 60208, United States
d.majerek@pollub.pl

Abstract. Water bodies often suffer from the discharge of nutrient loading from agricultural and urban areas that compromises the quality of water. This study presents the application of the Hotelling test to evaluate the impacts of treated wastewater, discharged from a municipal wastewater treatment plant (WWTP), on the quality of river water. The quality of water was described by different pollution indicators, including COD, BOD₅, TSS, NH₄-N, NO₂-N, NO₃-N, TKN, TN and TP. The water samples were collected at three different locations: 500 m above the discharge point, at the wastewater discharge point and 1000 m below the wastewater discharge point. The tests of single pollution indicator showed differences between the two locations. Specifically, the results show that each single comparison controlled type I error at 0.05, while the family-wise error rate for the tests of all marginal hypotheses was controlled at 0.37. Testing for single indicators separately may not reveal true multivariate differences. In order to overcome this limitation, a modified version of T² Hotelling test was used with robust James-Stein type estimators of covariance matrix. Major differences in the overall water quality were observed mainly for the concentration of nitrogenous compounds and found to significantly influence the water quality of the receiving river.

1. Introduction
Eutrophication of water bodies has become one of the major challenges of our society [1]. This challenge is represented by the discharge of detergents, fertilizers and/or sewage, usually rich in nutrients (e.g. nitrogen and phosphorus), in water systems. The oversupply of nutrients leads to an overgrowth of plants and algae in the aquatic systems; the death of these organisms and bacterial degradation of their biomass translates into hypoxia [2].

In the past, the most simple and convenient method for removing liquid and suspended solids involved discharging the waste into the receiving water bodies. The wastewater that was discharged into the rivers by means of gutters was diluted and underwent self-cleaning, eventually restoring its own quality within a certain distance away from the discharge site. In recent years, this method has become
obsolete due to the large amount of discharged wastewater and the impact of industrially synthesized chemicals causing both negative effects on the entire natural water cycle and on the human health [3,4]. As a result, only the wastewater that was subjected to appropriate treatment processes in a wastewater treatment plant (WWTP) would be discharged into the receiving water bodies (e.g. rivers), in accordance with the current local regulations [5-7].

The wastewater that has been effectively treated, can still be still a vector for a large number of conventional and non-conventional pollutants (e.g. micropollutants) that might not be eliminated by means of common methods. Additionally, on some occasions, a wastewater treatment failure can lead to the release of pollutants in the water bodies with subsequent eutrophication or raise of water salinity. Since these pollutants, along with others, can be transported by rivers over significant distances, it becomes even more crucial to establish a full set of parameters that allow for the assessment of the water river quality [8-10].

Physicochemical analyses have long been employed for the assessment of the running waters quality [11,12]. Basic parameters used to determine the wastewater and receiver waters quality include Chemical Oxygen Demand (COD), 5-day Biological Oxygen Demand (BOD), Total Suspended Solids (TSS), as well as measurements of phosphorus and nitrogen content. A river health assessment can also be thoroughly performed when the elements of biological and hydro-morphological assessment are introduced [13-15]. A simultaneous application of these assessment criteria can yield good results for monitoring the health status of rivers. However, this is seldom performed in practice [15]. Additionally, the majority of recent studies involve new evaluation techniques [5,15,16]. Very few studies have determined the sensitivity of the aforementioned parameters, and moreover they have not fully evaluated their agreement and complementarity with the overall health river status [17,18].

The main goal of this research is to explore the potential effects of treated wastewater discharge on the overall river quality. This was achieved by implementing the pollution indicators level, testing single variables separately and using omnibus $T^2$ Hotelling test.

2. Materials and methods

The water samples were collected at three different river locations, and the quality of the water was described by nine standard indicators that included COD, BOD$_s$, TSS, ammonia (NH$_4$-N), nitrite (NO$_2$-N), nitrate (NO$_3$-N), total Kjeldahl nitrogen (TKN), total nitrogen (TN), and total phosphorus (TP). In order to ensure the appropriate quality of the results, the sampling locations were at specific distances between each other: the first location was 500 m before the treated wastewater discharge point, second at the discharge point and the third location was 1000 m beyond the discharge point. The samples were collected from the municipal wastewater plant (South-Eastern Poland) once a week for three years. The data collected from the locations above and below the wastewater discharge were gathered once a month over one year. The samples were collected 50 cm below the water level by means of a bucket; these were subsequently poured into their containers following the Water and Wastewater Sampling Guidelines [19]. The nitrogen and phosphorus compounds were analysed using a HACH DR2800 spectrophotometer as well as the methods authorized by the HACH-Lange company (LCK 238 LATON, LCK 340, LCK 341, and LCK 303 for N; LCK 348 for P). The COD concentration was measured with the LCK 314 cuvette test, TSS were measured using the HACH DR2800 standard, and BOD$_s$ was analysed in accordance with the PN-EN 1899-1:2002 standard.

The main hypothesis of our research was the lack of differences in pollution indicators between the two locations above and below the discharge point, $H_0$: $Q_{\text{above}} = Q_{\text{below}}$ (null hypothesis). Given that only two points were considered, the discharge point was treated as “treatment” in the experiment. The level of concentration of the pollutants expressed by the value of indicators at the outlet of the municipal wastewater treatment plant was the potentially differentiating factor for the river water quality at the aforementioned locations. The pollution indicator readouts from above and below the discharge point were treated as dependent measurements, because they were obtained from two locations of the same river, slightly spaced apart [20].


All statistical calculations and visualizations were performed using the R environment [21]. The following R packages were used: dplyr, ggpubr, Hotelling, purrr, MVN and CovTestR [22-27].

3. Results
The tests of single pollutants indicator explain whether a particular indicator differs between two locations. Each single comparison controls type I error at a certain level, while the family-wise error rate for tests of all marginal hypotheses can vary [28]. Usually, the study of different pollutants does not take into account the relationships between variables. Hence, testing single dimension separately may not reveal true multivariate differences [29]. Controlling type I error in hypothesis testing at the 0.05 significance level means that the probability of rejection of the true null hypothesis would be less than 0.05 [30]. In order to overcome this limitation, omnibus tests like the $T^2$ Hotelling test are proposed as a generalization of Student’s t-test [31-33]. The covariance matrix, used in the calculation of $T_2$ statistic, is almost singular. Numerical issues make it challenging to carry out a classical $T^2$ Hotelling. The problem is in the formula of the test, where the statistics is the inverse of the covariance matrix, and there is a determinant close to zero, so that the values of the inverse take large values [34]. In this study, a modified version of $T^2$ Hotelling test, with robust James-Stein type estimators of covariance matrix, was used [35-37]. The robust estimator of the covariance matrix by James-Stein used a shrinkage method of obtaining estimators by reducing their variance. The assumptions were tested with the Mardia test [38], Box’s M test [39], Schott test [40], and permutation test [41].

The Hotelling’ test was applied with the following two assumptions: reliable multivariate normality and equality of covariance matrix of both populations. The first assumption was evaluated using the Mardia test [38]. Multivariate normality was evaluated by the Mardia test and it was performed in two steps: 1) the level of asymmetry was tested and 2) the kurtosis was calculated. If both statistics were similar to their counterparts for multidimensional normal distribution, then the test did not reject the hypothesis of multivariate normality. In our case, the test rejected the null hypothesis for kurtosis in both populations of the pollution indicator data. Figure 1 shows wide differences between groups in terms of kurtosis. The readouts from the location below the discharge of wastewater were characterized by a lower kurtosis for almost all pollution indicators. This translates into the observations from above the discharge point having a concentration around the mean higher than the ones below. This phenomenon, named heterokurticity, causes the rejection of multivariate normality of both phases.

The most common way of testing homogeneity of covariance is using the Box’s M test [39]. However, one of the limitations is that this only works for multidimensional normal distributions. Another test for heterokurtic populations (not normal), the Schott test, is a generalization of the Box’s M test [40]. This test rejects the null hypothesis; therefore, the $T^2$ Hotelling test with James-Stein robust covariance estimator could give unreliable results. Because of that, the hypotheses by the permutation test were also tested [41,42]. Permutation methods have been used, since they are a class of statistical tests that, under minimal assumptions, can provide exact control of false positives (i.e., type I error). The central assumption is simply the exchangeability that occurs by swapping different data points but keeping the data as the original. The permutation test was calculated based on several shuffled data labels of the grouping variable (number of permutations = 10000).

The relative differences in means ($Q_{below} - Q_{above})/Q_{above}$ of pollutant indicators between two locations was the greatest for NO$_2$-N and corresponded to 446%. The rest of nitrogenous compounds also had large relative differences varying from 36% to 105%. Single comparison tests for each pollution index controlled type I error at 0.05; however, the family-wise error rate for tests of all marginal hypotheses was 0.37 for the evaluated case. This value was rather high.

The Omnibus Hotelling test rejected the null hypothesis with $T^2=15.19$ and P-value = 9e-06 (p-value for permutation test was even lower). This confirms that there were differences in the water quality between two different locations. Univariate tests of particular pollutant indicators showed that only BOD$_5$, COD and TSS did not differ significantly. Figure 2 illustrates the observed differences.
Figure 1. Kernel densities of the concentrations of indicators with respect to measurement location: BOD, COD, TSS, NH$_4$-N, NO$_2$-N, NO$_3$-N, TKN, TN and TP.

Figure 2. Bar plots with error bars of mean concentration of pollutants above and below the treated wastewater discharge point into river: BOD, COD, TSS, NH$_4$-N, NO$_2$-N, NO$_3$-N, TKN, TN and TP. Whiskers indicates 95% confidence interval for the mean.
4. Conclusions

In our study we applied the $T^2$ Hotelling test to assess and compare the river water quality below and above the wastewater discharge point. The main differences in the water quality were observed in the case of nitrogenous compounds concentration; this significantly influenced the water quality of the receiving river.

Taking into account the relationship between different indicators is crucial when this comparison is carried out. Alternatively, there is a loss in the control of type I error of the test with subsequent unreliable results and in extreme cases, it becomes unfeasible to discriminate between groups of different indicators.

This article also shows that, given the lack of normality of distribution or singularity of covariance matrix, the classic form of the $T^2$ Hotelling test cannot be used; however, alternatively, a variety of generalization tests can be used.

The crucial methodological part of this paper is the application of multidimensional test for verification of the hypothesis. Bearing in mind the concern for methodological correctness, all assumptions were checked, and if they were not met, appropriate steps were taken.

References

[1] Paerl H W and Huisman J 2009 Climate change: a catalyst for global expansion of harmful cyanobacterial blooms Environ. Microbiol. Rep. 1 27–37
[2] Chislock M F, Doster E, Zitomer R A and Wilson A E 2013 Eutrophication: Causes, Consequences, and Controls in Aquatic Ecosystems Nature Education Knowledge 4 10
[3] Kundzewicz Z W, Zalewski M, Kędziora A and Pierzgalski E 2010 Zagrożenia związane z wodą Nauka 4 87–96
[4] Kujawskas A and Cel W 2017 Cluster analysis to evaluate chemical compounds in the waters of piezometers from a drilling wastes landfill ITM Web Conf. 15 04003
[5] Pliashechnyk V, Danko Y, Łagód G, Drewnowski J, Kuzmina T and Babko R 2018 Ciliated protozoa in the impact zone of the Uzhgorod treatment plant ed M Cimochowicz-Rybicka E3S Web Conf. 30 02008
[6] Czarnota J and Masłoń A 2019 Evaluation of the Effectiveness of a Wastewater Treatment Plant with MBBR Technology Rocz. Ochr. Srodowiska 906–925
[7] Dz.U. 2019 poz. 1311. Rozporządzenie Ministra Gospodarki Morskiej i Żeglugi Śródlądowej z dnia 12 lipca 2019 r.
[8] APHA 1992 Standard Methods for the Examination of Water and Wastewater. 18th edition. American Public Health Association (APHA), American Water Works Association (AWWA) and Water Pollution Control Federation (WPCF), Washington, DC
[9] Simeonov V, Stratis J A, Samara C, Zachariadis G, Voutsa D, Anthemidis A, Sofoniou M and Kouimtzis T 2003 Assessment of the surface water quality in Northern Greece Water Res. 37 4119–24
[10] Cieszynska M, Wesolowski M, Bartoszewicz M, Michalska M and Nowacki J 2012 Application of physicochemical data for water-quality assessment of watercourses in the Gdansk Municipality (South Baltic coast) Environ. Monit. Assess. 184 2017–29
[11] Hamaidi-Chergui F and Brahim Errahmani M 2019 Water quality and physicochemical parameters of outgoing waters in a pharmaceutical plant Appl. Water Sci. 9 165
[12] Srebotnjak T, Carr G, de Sherbinin A and Rickwood C 2012 A global Water Quality Index and hot-deck imputation of missing data Ecol. Indic. 17 108–19
[13] Babko R, Kuzmina T, Suchorab Z, Widomski M K and Franus M 2016 Influence of Treated Sewage Discharge on the Benthos Ciliate Assembly in the Lowland River Ecol. Chem. Eng. S 23 461–71
[15] Gorzel M, Kornijów R and Buczyńska E 2018 Quality of Rivers: Comparison of Hydro-Morphological, Physical-Chemical and Biological Methods Ecol. Chem. Eng. S 25 101–22
[16] Rajfur M and Kłos A 2015 Use of Algae in Active Biomonitoring of Surface Waters Ecol. Chem. Eng. S 21 561–76
[17] Mishra S, Kneis D, Berendonk T U and Aubeneau A 2019 Optimum positioning of wastewater treatment plants in a river network: A model-based approach to minimize microbial pollution Sci. Total Environ. 691 1310–9
[18] Arísti I, von Schiller D, Arroita M, Barceló D, Ponsati L, García-Galán M J, Sabater S, Elosegí A and Acuña V 2015 Mixed effects of effluents from a wastewater treatment plant on river ecosystem metabolism: subsidy or stress? Freshw. Biol. 60 1398–410
[19] PN-EN ISO/IEC 17025:2018
[20] Huang P, Ou A, Piantadosi S and Tan M 2014 Formulating appropriate statistical hypotheses for treatment comparison in clinical trial design and analysis Contemp. Clin. Trials 39 294–302
[21] R Core Team. 2019 R: A language and environment for statistical computing. www.r-project.org
[22] Wickham H, François R, Henry L, and Müller K 2019 Dplyr: A Grammar of Data Manipulation. https://CRAN.R-project.org/package=dplyr
[23] Kassambara A. 2019 Ggpubr: ‘Ggplot2’ Based Publication Ready Plots. https://CRAN.R-project.org/package=ggpubr.
[24] Curran J 2018 Hotelling: Hotelling’s T^2 Test and Variants. https://CRAN.R-project.org/package=Hotelling
[25] Wickham H 2019 Purrr: Functional Programming Tools. https://CRAN.R-project.org/package=purrr
[26] Korkmaz, Selcuk, Dincer Goksuluk, and Gokmen Zararsiz 2014 MVN: An R Package for Assessing Multivariate Normality The R Journal 6 151–62
[27] Barnard, Ben, and Dean Young 2018 CovTestR: Covariance Matrix Tests. https://CRAN.R-project.org/package=_covTestR
[28] Hochberg Y and Tamhane A C 1987 Multiple Comparison Procedures Hoboken, NJ, USA: John Wiley & Sons, Inc
[29] Coombs W T, Algina J and Oltman D O 1996 Univariate and Multivariate Omnibus Hypothesis Tests Selected to Control Type I Error Rates When Population Variances Are Not Necessarily Equal Rev. Educ. Res. 66 137–79
[30] Mudge J F, Baker L F, Edge C B and Houlanah J E 2012 Setting an Optimal That Minimizes Errors in Null Hypothesis Significance Tests ed Z Su PLoS One 7 e32734
[31] Hotelling H 1992 The Generalization of Student’s Ratio BT - Breakthroughs in Statistics: Foundations and Basic Theory ed S Kotz and N L Johnson (New York, NY: Springer New York) pp 54–65
[32] Perlman M D 2019 On the feasibility of Parsimonious variable selection for Hotelling’s T2-Test. (Preprint arXiv:1910.03699)
[33] Baringhaus L and Gaigall D 2017 Hotelling’s T2 Tests in Paired and Independent Survey Samples: An Efficiency Comparison J. Multivar. Anal. 154 177–98
[34] Wu Y, Genton M G and Stefanski L A 2006 A Multivariate Two-Sample Mean Test for Small Sample Size and Missing Data Biometrics 62 877–85
[35] Opgen-Rhein R and Strimmer K 2007 Accurate Ranking of Differentially Expressed Genes by a Distribution-Free Shrinkage Approach Stat. Appl. Genet. Mol. Biol. 6
[36] Schäfer J and Strimmer K 2005 A Shrinkage Approach to Large-Scale Covariance Matrix Estimation and Implications for Functional Genomics Stat. Appl. Genet. Mol. Biol. 4
[37] Tong T, Jang H and Wang Y 2012 James–Stein type estimators of variances J. Multivar. Anal. 107 232–43
[38] Mardia K V, Kent J T and Bibby J M 1979 Multivariate Analysis New York: Academic Press
[39] Box G E P 1949 A General Distribution Theory for a Class of Likelihood Criteria Biometrika 36 317–46
[40] Schott J 2001 Some tests for the equality of covariance matrices J. Stat. Plan. Inference 94 25–36
[41] Chung E and Romano J P 2016 Multivariate and multiple permutation tests J. Econom. 193 76–91
[42] Gamm B M and Gerard P 2020 Testing the differences between two color measurement probability distributions using Hotelling’s T2 test and the permutation test Color Res. Appl. 45 196–207