Journal of Biometry Studies

ISSN: 2791-7169
Year: 2024 - Volume: 4 - Issue: 1

PRENSIP

www.biometrystudies.com

Journal of Biometry Studies

Year: 2024, Volume: 4, Issue: 1

ISSN: 2791-7169

www.biometrystudies.com

www.prensip.gen.tr

Journal of Biometry Studies is an international peer-reviewed open access journal published biannually (June and December) in electronic form by *Prensip Publishing*. Submission and the publication process are free of charge.

Journal of Biometry Studies

Year 2024 – Volume 4 – Issue 1

www.biometrystudies.com

Journal of Biometry Studies

Year 2024 – Volume 4 – Issue 1

CONTENTS

Research Articles

RESEARCH ARTICLE
Journal of Biometry Studies (2024) 4(1): 1-10
DOI: 10.61326/jofbs.v4i1.01
www.biometrystudies.com

Journal of Biometry Studies
PRENSIP
www.prensip.gen.tr

Exploring ecosystem health through physico-chemical parameters and bioindicators in Miagao, Iloilo, Philippines: A preliminary study

Maria Liza B. TORING-FARQUERABAO[1,2], Angelica B. BERMIL[1], Albaris B. TAHILUDDIN[1,3,*]

[1]Mindanao State University-Tawi-Tawi College of Technology and Oceanography, College of Fisheries, Sanga-Sanga, Bongao, Tawi-Tawi/PHILIPPINES
[2]University of the Philippines Visayas, College of Fisheries and Ocean Sciences, Institute of Marine Fisheries and Oceanology, Miagao, Iloilo/PHILIPPINES
[3]Kastamonu University, Institute of Science, Department of Aquaculture, Kastamonu/TÜRKİYE

[*]Corresponding author: albaristahiluddin@msutawi-tawi.edu.ph Received: 19/01/2024, Accepted: 22/06/2024

Abstract

Ecological health surveys are crucial for evaluating the well-being of ecosystems, as they provide essential insights through surrogate indicators, enabling informed decision-making for sustainable environmental management and conservation. This study conducted an ecosystem health survey in Miagao, Iloilo, Philippines, with specific objectives including the evaluation of physico-chemical parameters, identification of plankton at the family level, microbiological analysis, and detection of photosynthetic bacteria in selected sites. Four sites were chosen, and physico-chemical parameters in sites 1 and 4 were almost identical. Plankton presence was observed, but quantification was not conducted. Microbiological analysis in Site 1 indicated that the total plate count was 10^5 CFU mL^{-1}. Growth of photosynthetic bacteria was observed in samples with sediment after 14 days. Despite limitations in site coverage and real-time insight, the study highlights the presence of bioindicators, offering valuable groundwork for future research and the need for comprehensive analysis across all sites.

Keywords: Ecosystem, Ecosystem health survey, Microbiological analysis, PSB

Please cite this article as follows:

Toring-Farquerabao, M. L. B., Bermil, A. B., & Tahiluddin, A. B. (2024). Exploring Ecosystem Health through Physico-chemical Parameters and Bioindicators in Miagao, Iloilo, Philippines: A Preliminary Study. *Journal of Biometry Studies*, *4*(1), 1-10. https://doi.org/10.61326/jofbs.v4i1.01

1. Introduction

Ecological health can be viewed in terms of ecosystems, where structural and functional characteristics are maintained (Manickavasagam et al., 2019). As ecosystem health cannot be measured or observed directly, surrogate measures (indicators) must be applied to assess it (Burkhard et al., 2008). Each organic entity inside a biological system provides an indication of the health of its surroundings (Parmar et al., 2016). One functional way to assess the health of ecosystems is to measure or model indicators that represent important processes in the ecosystem, from which conclusions for the–whole system can be drawn (Burkhard et al., 2008).

For instance, organisms known as bioindicators–living entities such as plants, plankton, animals, and microbes–are utilized to monitor the health of the natural ecosystem (Parmar et al., 2016). Bioindicators indicate altered environmental conditions and can be used to identify and/or quantify the impact of pollutants on the environment. Assessing the status of bioindicators indirectly provides an estimate of the natural state or the level/degree of contamination present in that particular ecosystem. It can be used as an index of measures or a model that characterizes ecosystem health (Manickavasagam et al., 2019). Bioindication and biomonitoring have become promising methods for studying the impacts of external factors on an ecosystem, its development, and for differentiating polluted and unpolluted areas. The environment's condition is effectively monitored by using bioindicator species due to their resistance to ecological variability (Parmar et al., 2016).

The Municipality of Miagao, a province in the Philippines, is a coastal area located forty kilometers from Iloilo City, along the southwestern part of the province of Iloilo, Panay Island. It has 22 coastal barangays out of the 119 barangays, with a 16 km coastline covering 24,000 hectares of municipal waters with rich fishery resources. A significant portion of its populace relies on coastal and marine resources for subsistence and livelihood. In fact, the 21,896 coastal population, comprising 34% of the 64,545 total population (NSO, 2010; Pilapil-Añasco et al., 2016), directly and indirectly depends on fishing as their source of livelihood. As of 2015, the Municipal Fisherfolk Registration (FishR) Program of BFAR 6 and LGU Miagao has a total of 2,360 registered fisherfolks, about four percent (4%) of the total population. According to the study of Chen et al. (2019a), the utilization of water resources has increased, and aquatic ecosystems have been seriously degraded or destroyed in recent years. The heights of dependability of the locals on its coastal resources made this survey information crucial to the locality. The study will serve as a preliminary report on the current status of the ecosystem's health in the area. Hence, this study aims to assess the ecosystem health status of selected sites in Miagao, Iloilo, Philippines. Specifically, this study aims to: (1) evaluate the physico-chemical parameters of the sampling areas; (2) identify, to the family level, the presence of plankton in two selected sites; (3) determine the bacterial count of a selected site; and (4) determine the presence of photosynthetic bacteria in selected sites using different improvised culture media.

2. Material and Methods

2.1. Study site

The study was conducted in Miagao, Iloilo, Philippines. Four specific sites were chosen for this study, namely: Site 1 – coast near the University of the Philippines (UPV) hatchery; Site 2 – coast near the UPV Ocean Weather Laboratory (OWL) building; Site 3 – waters near Sulu Restaurant; and Site 4 – coast near Barangay Sapa (Figure 1). The choosing of the sites was based on the available potential domestic wastes that could possibly alter the ecosystem's health. Site 1 is closer to hatchery site where wastes were directly discharged into the water. Site 2 and 3 are near residential areas while Site 4 is along a commercial restaurant. In addition, the sites chosen were close to each other due to some logistical constraints such as time and availability of materials in the laboratory. However, through the collection of water parameters of the chosen sites, these may have a group of organisms that establish bioindicators that can indicate the health of the ecosystem and have an impact on the community that relies on fisheries. Furthermore, each experiment only pertains to specific sites and does not necessarily cover all four mentioned areas. All analyses were conducted at the UPV laboratory.

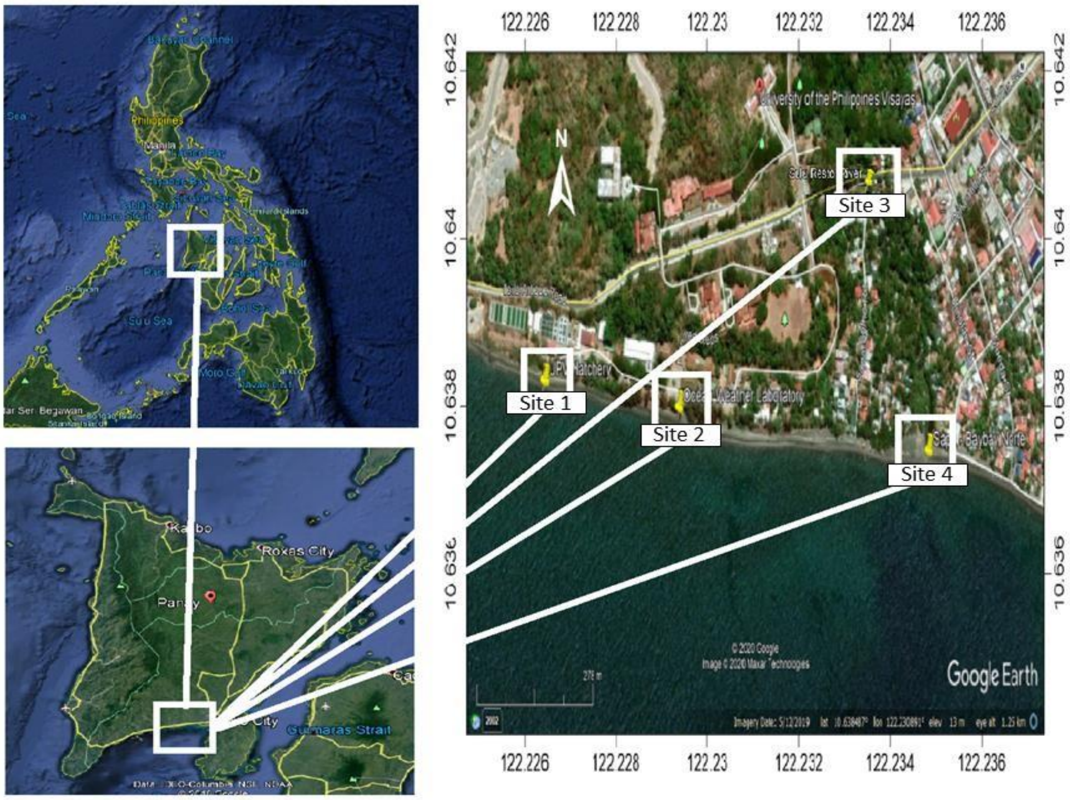

Figure 1. Map of the study site

2.2. Evaluation of physico-chemical parameters and plankton identification

Physico-chemical parameters such as salinity, temperature, and total dissolved solids (TDS) were recorded using a YSI Pro Plus multi-parameter water quality meter. Due to time constraints and laboratory resource limitations, only two sites were sampled: the coast near the UPV hatchery (Site 1) where hatchery discharges are located and Barangay Sapa (Site 4), situated closer to coastal residences. These sites were chosen to assess bioindicators indicative of ecosystem health. Additionally, the type of substrate and weather conditions at each sampling location were noted for plankton identification. A plankton tow net with a 50 µm mesh size, 12.0 cm diameter, and 30 cm length was used to collect water samples. The net was manually dragged along the coast, and the collected samples were then transferred to separate, properly labeled plastic containers for each site. These containers were transported to the IMFO-UPV laboratory for identification using a compound microscope. Samples were identified up to the family level using the identification manual (Verlecar & Desai, 2004) and subsequently photographed for documentation and analysis purposes. A one-time sampling event was conducted as this study aims solely to provide a preliminary assessment of the chosen sites' ecosystem health.

2.3. Glasswares, equipment and media preparation

All glassware, including test tubes, petri dishes, L-rods, pipettes, pipette tips, etc., were thoroughly washed and sterilized to prevent any contamination. The sterilization process took place at the IMFO-UPV Laboratory. On the other hand, media preparation was conducted at the Institute of Aquaculture (IA), UPV. The media used are listed in Table 1. These media were accurately weighed with their corresponding concentrations and then autoclaved. Subsequently, they were poured into petri dishes and properly labeled.

Table 1. Media used in bacterial counting in this study

Treatment No.	Media Used
1	nutrient agar (NA) + filtered seawater (FSW)
2	tryptone + yeast extract + agar-agar + FSW
3	NA + 2% NaCl + distilled water

2.4. Microbiological analysis

A water sample from one site, specifically the coast near UPV hatchery, was taken to the Microbiology Laboratory of IA-UPV for the total plate count (TPC) of bacteria present. Only one site was processed instead of two due to time constraints and the availability of materials in the laboratory. Five-fold dilutions were prepared with sterile saline solution. An aliquot (100 µL) of each sample dilution was spread onto plates with prepared media using an L-rod. The plates were then inverted, placed inside transparent plastic cellophane, and stored in the laboratory at room temperature for 72 hours. Only the first, third, and fifth dilutions were used in this study. Two replicates per sample were prepared and executed. All plates were monitored every 24 hours, and the number of colonies present was counted. For the calculation of bacteria colonies using the TPC method, the following procedure was employed:

$$\text{TPC} = \frac{\text{Number of colonies}}{\text{Dilution Factor x Volume Plated}}$$

2.5. Determination of the Presence of Photosynthetic Bacteria (PSB)

Media preparation

An egg solution was served as fertilizer and used to aid the growth of PSB in this experiment. The ingredients used were 2 eggs, 2 tablespoons of monosodium glutamate (MSG), and 1 tablespoon of soy sauce. These were mixed thoroughly and added to each water sample collected from each study site.

Culture method and monitoring of PSB growth

Collected water and sediment samples from all four sites were used in this experiment. Transparent plastic bottles with a volume of 200 mL were used, and a total of ten treatments were prepared (Table 2). The media, i.e., the fertilizer aiding in bacterial growth, was then added to each bottle. Each site had two treatments: water samples only and water together with its sediments. The sediment for the control treatment was gathered from terrestrial soil nearby. Ten bottles were prepared, appropriately labeled, and stored at the Faculty Center Building of UPV, where sufficient sunlight is available (Figure 2). Monitoring was conducted once a week for 21 days. Each bottle was thoroughly observed for any progress, and observations were noted. Samples were also photographed during each visit for documentation and analysis purposes.

Table 2. Treatments used in PSB culture in this study

Treatment No.	Samples Used
1	coast near hatchery of UPV; water only
2	coast near hatchery of UPV; water + sediment
3	coast near OWL building of UPV; water only
4	coast near OWL building of UPV; water + sediment
5	waters near Sulu Restaurant; water only
6	waters near Sulu Restaurant; water + sediment
7	coast near Barangay Sapa; water only
8	coast near Barangay Sapa; water + sediment
9	distilled water (control)
10	distilled water + soil (control)

Figure 2. Experimental samples used in PSB culture

3. Results and Discussion

3.1. Physico-chemical parameters

Physico-chemical and biological characteristics dictate the aquatic ecosystem's health (Venkatesharaju et al., 2010). According to Pal and Chakraborty (2014), good quality of water resources depends on a large number of physico-chemical parameters and biological characteristics. Water quality is an important component of a water ecosystem health assessment (Chen et al., 2019a). In this study, only two sites, namely sites 1 and 4, were considered in this part of the experiment, as these two sites were initially planned for microbiological analysis. The selection of sites was limited due to the availability of materials in the laboratory. The water sample in site 1 was collected from an area near hatchery discharges, while the water sample in site 4 was collected from an area near coastal community residences. Physico-chemical parameters for the two sites did not show any significant differences since the two locations are less than a kilometer away (Table 3). The only distinction between the two is the type of substrate each area has, i.e., sandy for site 1 and rocky for site 4. Water parameters were measured during a sunny day; hence, the temperature is quite high, coinciding with its salinity.

Table 3. Physico-chemical parameters

Sampling Site	Type of substrate	Salinity (ppt)	Temperature (°C)	Total dissolved solids (TDS)
1	Sandy	32.7	27	32.06
4	Rocky	33.0	27	32.27

The first bioindicators identified in this study are plankton. In many water bodies, such as seas, lakes, streams, and swamps, significant biological production is carried out by plankton. These organisms consist of communities that float along currents and tides, yet they fuse and cycle important quantities of energy that are then passed on to higher trophic levels (Walsh, 1978). The presence of plankton is key to marine organisms, serving as both an indicator of water quality and the main food source for many fish (Thakur et al., 2013). The changes that occur within the communities of plankton provide the platform to determine the trophic state of water bodies (Pradhan et al., 2008). These organisms respond rapidly to changes taking place in the surrounding environment and serve as important biomarkers for assessing the quality of water as well as indicators of water pollution (Parmar et al., 2016).

Plankton includes both plants, referred to as phytoplankton, and animals referred to as zooplankton. As shown in Table 4, both sites have the presence of diatoms and dinoflagellates, which are phytoplankton. Phytoplankton have been used for the successful observation of water contamination and are a useful indicator of water quality (Wu, 1984). Changes in the diversity of phytoplankton species may indicate pollution of the marine ecosystem (Walsh, 1978; Hosmani, 2014). However, this study only focuses on the identification of plankton present at the family level. Counting was not done in this study due to some unexpected circumstances (i.e., time due to the COVID-19 pandemic) where prohibitions were enforced and all samples left in the laboratory were left out and damaged; hence, conclusions about the status of the ecosystem in the locality are not possible. In the context of altering nutrient charges in aquatic systems, the ratio of the major phytoplankton groups, diatoms versus flagellates (diatoms–nondiatoms ratio), can be used as an indicator. For example, regarding eutrophication, nutrient reductions can be observed in a decrease in flagellate abundance (Burkhard et al., 2008). Diatoms are powerful indicators of environmental change and have emerged as preferred indicators in monitoring studies (Dixit et al., 1992). In the study of Yusuf (2020), the presence of organic pollution indicators *Closterium* sp., *Navicula* sp., *Nitzschia* sp., *Synedra* sp., *Chlamydomonas* sp., *Cyclotella* sp,. and *Anacystis* sp. is a warning sign of the deteriorating condition of the water quality in the reservoir. Therefore, quantifying the presence of plankton and identifying it at the species level will give more weight to overseeing the real state of the ecosystem.

For the zooplankton identified in the study, both sites exhibited the same presence of copepods. Identification of the zooplankton plays a vital role and forms the intermediate trophic status between phytoplankton communities and fish groups which serves as a necessary component of the aquatic ecosystem (Pal & Chakraborty, 2014). The organism plays a vital role in the food chain, nutrient recycling, and energy flow in the aquatic ecosystem (Park et al., 2007). According to Aslam et al. (2012), copepods (cyclops & phyllodiaptomus) indicate the health of the marine body. These organisms are microscopic animals living near the surface of the water body. They are poor swimmers, instead relying on tides and currents as a transport mechanism. Zooplankton also play an important role as bioindicators and help to evaluate the level of water pollution (Parmar et al., 2016). These types of organisms play a pivotal role in aquatic food webs by transferring carbon to higher trophic levels, consuming microorganisms (bacteria, protists), and serving as prey for fish and invertebrates (Drira et al., 2018). The presence or absence of certain zooplankton species may indicate the relative influence of different water types on ecosystem structures and may serve as an early indication of a biological response to environmental and climatic changes (Hays et al., 2005; Ziadi et al., 2015). They are identified as excellent bioindicators to evaluate the contamination of any coastal and oceanic bodies (Zannatul & Muktadir, 2009; Parmar et al., 2016). Quantification of these organisms is of great importance as well to have an overview of the area's ecosystem health. However, determination of their abundance was not done in this experiment.

Table 4. Plankton identified in two study sites

Sampling Site	Phytoplankton	Family	Zooplankton
1	Diatoms	Rhizosoleniaceae Coscinodiscaceae Thallasionemataceae Leptocylindraceae Naviculaceae Chaetocerotacea Bacillariophyceae	Copepods
	Dinoflagellate	Pyrocyctis Ceratium Gymodinium	
2	Diatoms	Bacillariophyceae Rhizosoleniaceae Coscinodiscaceae Lithodesmiaceae Chaetocerotacea Eupodiscacea	Copepods
	Dinoflagellate	Ceratium	Fish Larvae Copepods Zoea

3.2. Microbiological analysis

Microorganisms play a crucial role in maintaining and sustaining any ecosystem, as they are more capable of rapid adjustment toward environmental changes and deterioration (Sorokin, 1981; Dash et al., 2012). Marine microbial communities form the basis of the ocean food web and, therefore, produce food for all life in the ocean (Glöckner et al., 2012). They can also serve as indicators of aquatic or terrestrial ecosystem health. Due to their abundance, they are easy to test and readily available (Parmar et al., 2016). One such marine microorganism is bacteria (Glöckner et al., 2012), which is used in this study to evaluate the ecosystem's health status of a selected site in Miagao, Iloilo. The site near UPV hatchery was chosen due to its location where discharges from the said facility are the focus of water sampling. Results revealed that in almost all of the tested media, abundances were consistently at almost the same level, approximately 10^5 CFU mL^{-1}, indicating a higher bacterial presence at the sampling site, likely due to its proximity to the hatchery system where effluents are discharged. Bioindicators are evidently present in this area, suggesting a lack of bioremediation. Although the bacteria are not identified, it can be inferred that the sampling area is considered biologically polluted. Microbial pollution of coastal areas can harm aquatic wildlife, i.e., potentially detrimental to coral reef community health, as well as human health, leading to many skin and intestinal diseases, thus reducing the benefits that coastal environments provide to the community (Basili et al., 2021; Ochsenkühn et al., 2021).

Table 5. Total plate count (TPC, CFU mL^{-1}) of bacteria from UPV hatchery water sample

Treatment	Incubation period (h)		
	24	48	72
1 (NA+FSW)	3.0×10^5	2.0×10^5	2.0×10^5
2 (tryptone+yeast+ NaCl+agar-agar+FSW)	3.0×10^3	2.0×10^5	5.0×10^5
3 (NA+2%Nacl+ distilled water)	2.6×10^6	4.0×10^5	8.5×10^5

3.3. Growth of photosynthetic bacteria

Bacteria have long been known to play a part in marine ecosystems (Sorokin, 1981). A particular strain of these microorganisms is known as photosynthetic bacteria (PSB). PSB are the earliest prokaryotes with a primitive photo energy synthesis system on earth (Chen et al., 2020) and are widely distributed in oceans, lakes, soil, and activated sludge (Zhou et al., 2015), possessing light-absorbing pigments (Talaiekhozani & Rezania, 2017). This means that this type of bacteria harvests light energy using specialized pigments (the 'photo' part) and can convert CO_2 into organic carbon (the 'synthesis' part), although with differing efficiencies (Karl, 2002). There are four families of PSB: Rhodospirillaceae (purple non-sulfur bacteria, PNSB), Chromatiaceae (purple sulfur bacteria), Chlorobiaceae (green sulfur bacteria), and Chloroflexaceae (gliding filamentous green sulfur bacteria) (Blankenship et al., 1995). These bacteria have versatile metabolic pathways (Madukasi & Zhang, 2010). Under light conditions, their metabolic pathways are photoautotrophic or photoheterotrophic; whereas under dark conditions, the metabolic pathway is chemo-heterotrophic (Larimer et al., 2004) hence enabling them to survive under light-anaerobic or dark-aerobic conditions (Talaiekhozani & Rezania, 2017). Such flexibility enables PSB to utilize various types of substrates, including organics, N, P, and S (He et al., 2010).

PSB cells are rich in high-valued substances, in which the protein content is usually 45%–65% (Yang et al., 2018). These high-value substances can be utilized as animal feed, fertilizer, agents of disease prevention, food and cosmetic supplements, and soil conditioners (Chen et al., 2019b). The photosynthetic bacteria also offer numerous advantages for the bioremediation process due to their capacity to utilize various types of organic compounds. They have the ability to remove COD, nitrogen, phosphorous, and different types of heavy metals from wastewater (Talaiekhozani & Rezania, 2017). Moreover, in the study of Chen et al. (2019b), PSB are used in wastewater treatment, i.e., a new technology that can simultaneously remove pollutants from wastewater and produce high-value substances. Numerous studies conducted by different authors (e.g., Kaewsuk et al., 2010, Chitapornpan et al., 2013, Prachanurak et al., 2014, Zhou et al., 2014) showed and noted that PSB have been utilized for the treatment of different wastewaters from sugar industries, food processing, chicken abattoirs, dairy, and fermented starch. This type of bacteria is resistant to salty environments; therefore, they have good potential for treating different high-organics-load wastewaters (Qin et al., 2017). Furthermore, PSB are not only applicable for wastewater treatment but are also used to generate quality-added products such as biodiesel and hydrogen (Talaiekhozani & Rezania, 2017). Hence, PSB serve a significant purpose in the marine environment.

In this study, the culturing of PSB utilizing an egg solution as fertilizer in water samples gathered from each selected site was used to determine the presence of these organisms. Observed growth of PSB, as shown in Figure 3, was noticed after 14 days of culture. The areas where these PSB grow are water samples gathered from the coast near the UPV hatchery, UPV-OWL building, and Barangay Sapa. In addition to that, samples with sediments are the ones where PSB successfully grow. The purple colors in each bottle designate the presence of this type of bacteria. According to Kolber et al. (2000), PSB has several independent lines of purple pigments which this experiment is based on. The presence of purple carotenoids such as spirilloxanthin gives these bacteria their distinct color (Jagannathan & Golbeck, 2009). The rest of the water samples do not exhibit any growth of PSB after 21 days of culture. Fleischman (2011) stated that PSB plays many important roles in the environment, as much as a third of the earth’s photosynthesis is performed by microorganisms in the oceans. The areas where PSB were being observed are all water samples gathered from a coast hence coinciding with the studies that these organisms can be found in marine waters. In the study of Orel et al. (2022), these types of bacteria have a high ability to remove heavy metals, dyes, and macro-pollutants from wastewater and a promising technology to treat different types of wastewater effectively and economically. As for the control sample, it showed no signs of growth, and the other site, i.e., near Sulu Restaurant, is subjected to some domestic discharges which may have affected the salinity of the area. However, further investigation must be done since water parameters were not being gathered during the experiment.

The results of the study would provide the local government and community with an overview of the ecosystem health status of the locality. Without clear definitions of ecosystem health, it is not possible to set targets and assess whether management actions have been effective. Bioindicators have remarkable potential in forecasting disasters, preventing pollution, and exploring and conserving natural resources, all aiming at sustainable development with minimum destruction of the biosphere (Manickavasagam et al., 2019). The findings of this study will serve as baseline information for future coastal resource management planning by responsible agencies and will also serve as a trial experiment. It can be applied to predict the impact of anthropogenic activities, particularly pollutants, and predict environmental changes in a timely manner. Informing the public about the real-time status of their coastal water is of high importance. Even though the concept of ecosystem health has been criticized for being too fuzzy and not concrete enough for practical application, Burkhard et al. (2008) stress that it still found entry into different management strategies and definitions of political targets. Furthermore, O’Brien et al. (2016) emphasize that assessing ecosystem health is an ongoing priority for governments, scientists, and managers worldwide.

4. Conclusion

In conclusion, with the preliminary data gathered, ecosystem health surveys in coastal areas, particularly in regions like Miagao, Iloilo, Philippines, are crucial for communities heavily dependent on fisheries. Additionally, time constraints and limited equipment in the laboratory are likely to mean that not all potential variables can be monitored. Hence, this paper highly recommends evaluating the sites thoroughly again for comprehensive data. Ecosystem health, indicative of sustainability and minimal external support, is intertwined with the well-being of both the environment and its human inhabitants. The assessment's results offer real-time insights into the coastal ecosystem, enabling informed decision-making for the benefit of both humans and the environment. Moreover, the preliminary data can help the policymakers/managers to make appropriate decisions to gather a more comprehensive scientific record for the better management of the coastal ecosystems studied.

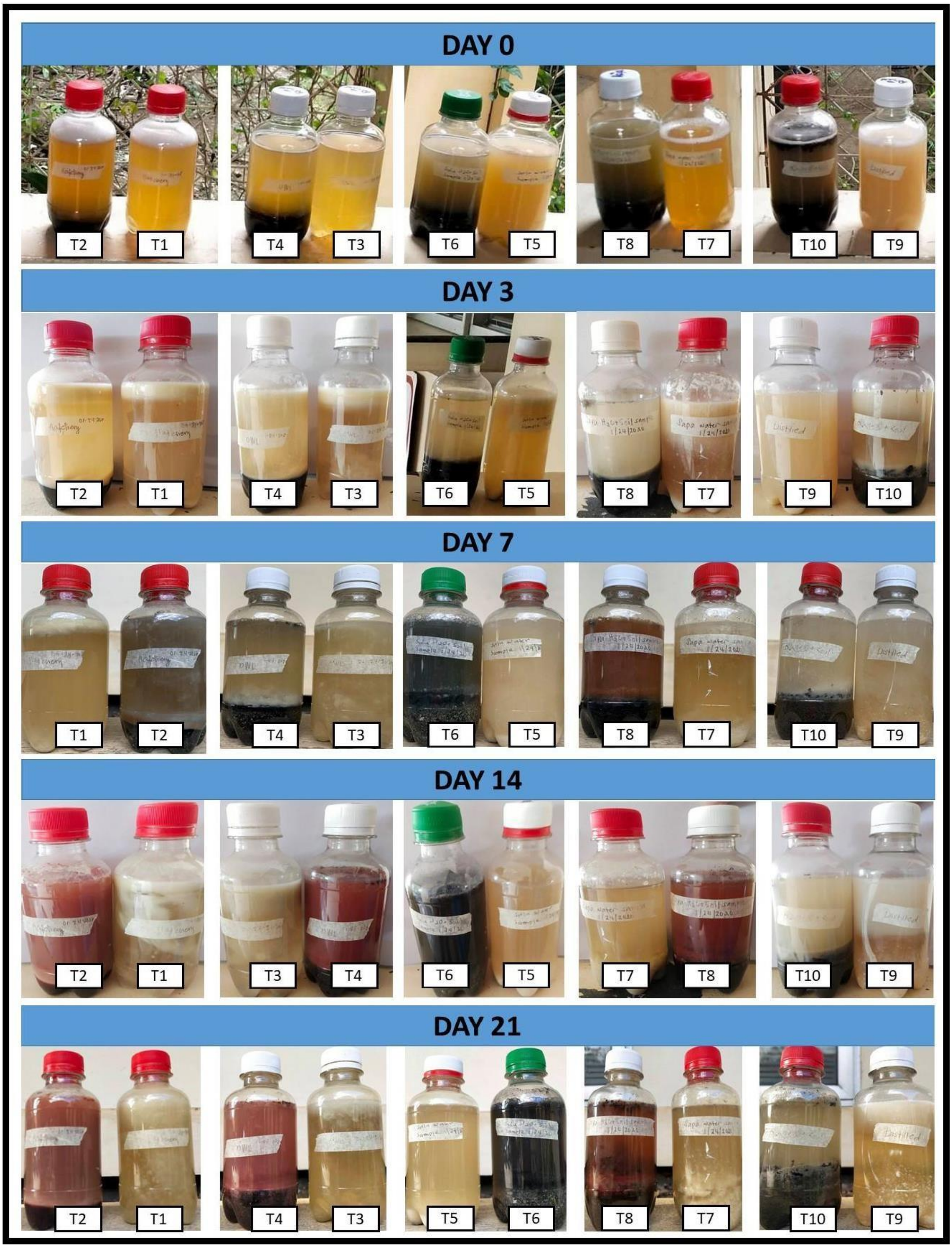

Figure 3. Monitoring of PSB growth for 21 days in this study

Conflict of interest

The authors declare that there is no conflict of interest.

Ethical Approval

For this type of study, formal consent is not required.

References

Aslam, M., Verma, D. K., Dhakerya, R., Rais, S., Alam, M., & Ansari, F. A. (2012). Bioindicator: A Comparative Study on Uptake and Accumulation of Heavy Metals in Some Plant`s Leaves of M.G. Road, Agra City, India. *Research Journal of Environmental Earth Sciences*, *4*(12), 1060-1070.

Basili, M., Campanelli, A., Frapiccini, E., Luna, G. M., & Quero, G. M. (2021). Occurrence and Distribution of Microbial Pollutants in Coastal Areas of the Adriatic Sea Influenced by River Discharge. *Environmental Pollution*, 285, 117672. https://doi.org/10.1016/j.envpol.2021.117672

Blankenship, R. E., Madigan, M. T., & Bauer, C. E. (1995). *Anoxygenic photosynthetic bacteria*. Kluwer Academic Publishers.

Burkhard, B., Müller, F., & Lill, A. (2008). Ecosystem Health Indicators. In JS. E. Jørgensen, & B. D. Fath (Eds.), *Encyclopedia of Ecology* (Vol. 2, pp. 1132-1138). Elsevier. https://doi.org/10.1016/B978-008045405-4.00312-8

Chen, J., Wang, Y., Li, F., & Liu, Z. (2019a). Aquatic Ecosystem Health Assessment of a Typical Sub-basin of the Liao River based on Entropy Weights and a Fuzzy Comprehensive Evaluation Method. *Scientific Reports*, *9*, 14045. https://doi.org/10.1038/s41598-019-50499-0

Chen, Y., Yang, A., Meng, F., & Zhang, G. (2019b). Additives for Photosynthetic Bacteria Wastewater Treatment: Latest Developments and Future Prospects. *Bioresource Technology Reports*, *7*, 100229. https://doi.org/10.1016/j.biteb.2019.100229

Chen, J., Wei, J., Ma, C., Yang, Z., Li, Z., Yang, X., Wang, M., Zhang, H., Hu, J., & Zhang, C. (2020). Photosynthetic Bacteria Based Technology is a Potential Alternative to Meet Sustainable Wastewater Treatment Requirement? *Environment International*, *137*, 105417. https://doi.org/10.1016/j.envint.2019.105417

Chitapornpan, S., Chiemchaisri, C., Chiemchaisri, W., Honda, R., & Yamamoto, K. (2013). Organic Carbon Recovery and Photosynthetic Bacteria Population in an Anaerobic Membrane Photo-bioreactor Treating Food Processing Wastewater. *Bioresource Technology*, *141*, 65-74. https://doi.org/10.1016/j.biortech.2013.02.048

Dash, H. R., Mangwani, N., Chakraborty, J., Kumari, S., & Das, S. (2012). Marine Bacteria: Potential Candidates for Enhanced Bioremediation. *Applied Microbiology and Biotechnology*, *97*(2), 561-571. https://doi.org/10.1007/s00253-012-4584-0

Dixit, S. S., Smol, J. P., Kingston, J. C., & Charles, D. F. (1992). Diatoms: Powerful Indicators of Environmental Change. *Environmental Science & Technology*, *26*(1), 22-33. https://doi.org/10.1021/es00025a002

Drira, Z., Kmiha-Megdiche, S., Sahnoun, H., Tedetti, M., Pagano, M., & Ayadi, H. (2018). Copepod Assemblages as a Bioindicator of Environmental Quality in Three Coastal Areas under Contrasted Anthropogenic Inputs (Gulf of Gabes, Tunisia). *Journal of the Marine Biological Association of the United Kingdom*, *98*(8), 1889-1905. https://doi.org/10.1017/S0025315417001515

Fleischman, D. (2011). Photosynthesis. In N. Sperelakis (Ed.), *Cell Physiology Source Book (4th Ed.)* (pp. 909-924). Academic Press. https://doi.org/10.1016/B978-0-12-387738-3.00051-2

Glöckner, F. O., Gasol, J. M., McDonough, N., & Calewaert, J. (2012). *Marine microbial diversity and its role in ecosystem functioning and environmental change*. Marine Board-European Science Foundation.

Hays, G. C., Richardson, A. J., & Robinson, C. (2005). Climate Change and Marine Plankton. *Trends in Ecology and Evolution*, *20*(6), 337-344. https://doi.org/10.1016/j.tree.2005.03.004

He, J., Zhang, G., & Lu, H. (2010). Treatment of Soybean Wastewater by a Wild Strain *Rhodobacter sphaeroides* and to Produce Protein under Natural Conditions. *Frontiers in Environmental Science & Engineering in China*, *4*, 334-339. https://doi.org/10.1007/s11783-010-0239-5

Hosmani, S. (2014). Freshwater Plankton Ecology: A Review. *Abhinav-International Monthly Refereed Journal of Research in Management and Technology*, *3*, 1-10.

Jagannathan, B., & Golbeck, J. H. (2009). Photosynthesis: Microbial. In M. Schaechter (Ed.), *Encyclopedia of Microbiology (3rd ed.)* (pp. 325-341). Elsevier. https://doi.org/10.1016/B978012373944-5.00352-7

Kaewsuk, J., Thorasampan, W., Thanuttamavong, M., & Seo, G. T. (2010). Kinetic Development and Evaluation of Membrane Sequencing Batch Reactor (MSBR) with Mixed Cultures Photosynthetic Bacteria for Dairy Wastewater Treatment. *Journal of Environmental Management*, *91*(5), 1161-1168. https://doi.org/10.1016/j.jenvman.2010.01.012

Karl, D. M. (2002). Hidden in a Sea of Microbes. *Nature*, *415*, 590-591. https://doi.org/10.1038/415590b

Kolber, Z. S., Van Dover, C. L., Niederman, R. A., & Falkowsk, P. G. (2000). Bacterial Photosynthesis in Surface Waters of the Open Ocean. *Nature*, *407*, 177-179. https://doi.org/10.1038/35025044

Larimer, F. W., Chain, P., Hauser, L., Lamerdin, J., Malfatti, S., Do, L., Land, M. L., Pelletier, D. A., Beatty, J. T., Lang, A. S., Tabita, F. R., Gibson, J. L., Hanson, T. E., Bobst, C., Torres, J. L., Peres, C., Harrison, F. H., Gibson, J., & Harwood, C. S. (2004). Complete Genome Sequence of the Metabolically Versatile Photosynthetic Bacterium *Rhodopseudomonas palustris*. *Nature Biotechnology*, *22*(1), 55–61. https://doi.org/10.1038/nbt923

Madukasi, E. I., & Zhang, G. (2010). Microaerobic Biodegradation of High Organic Load Wastewater by Phototrophic Bacteria. *Journal of Biotechnology*, *9*, 3852-3860. https://doi.org/10.5897/AJB10.186

Manickavasagam, S., Sudhan, C., Bharathi, & Aanand, S. (2019). Bioindicators in Aquatic Environment and their Significance. *Journal of Aquatic Tropics*, *34*(1), 73-79. http://doi.org/10.32381/JAT.2019.34.01.6

NSO (2010). Census of population and housing. https://psa.gov.ph/content/2010-census-population-and-housing-reveals-philippine-population-9234-million

O'Brien, A., Townsend, K., Hale, R., Sharley, D., & Pettigrove, V. (2016). How is Ecosystem Health Defined and Measured? A Critical Review of Freshwater and Estuarine Studies. *Ecological Indicators*, *69*, 722-729. http://dx.doi.org/10.1016/j.ecolind.2016.05.004

Ochsenkühn, M.A., Fei, C., Bayaara, O., Romeo, E., Amosa, P., Idaghdour, Y., Goldstein, G., Bromage, T.G., & Amin, S.A. (2021). Microbial Contamination Survey of Environmental Fresh and Saltwater Resources of Upolu Island, Samoa. *Environments 8*(11), 112. https://doi.org/10.3390/environments8110112

Orel, N., Fadeev, E., Klun, K., Licer, M., Tinta, T., & Turk, V. (2022) Bacterial Indicators Are Ubiquitous Members of Pelagic Microbiome in Anthropogenically Impacted Coastal Ecosystem. *Frontiers in Microbiology*, *12*, 765091. https://doi.org/10.3389/fmicb.2021.765091

Pal, S., & Chakraborty, K. (2014). Importance of Some Physical and Chemical Characteristics of Water Bodies in Relation to the Incidence of Zooplanktons: A Review. *Indian Journal of Social and Natural Sciences*, *3*, 102-116.

Park, K. S., & Shin, H. W. (2007). Studies on Phyto-and-Zooplankton Composition and Its Relation to Fish Productivity in a West Coast Fish Pond Ecosystem. *Journal of Environmental Biology*, *28*(2), 415-422.

Parmar, T. K., Rawtani, D., & Agrawal, Y. K. (2016). Bioindicators: The Natural Indicator of Environmental Pollution. *Frontiers in Life Science*, *9*(2), 110-118. https://doi.org/10.1080/21553769.2016.1162753

Pilapil-Añasco, C., Baylon, C., Nequia, E. L., & Gulbranson, J. (2016). Integrated Coastal Management Plan (ICMP) 2016-2026 Miagao, Iloilo. 81 pp.

Prachanurak, P., Chiemchaisri, C., Chiemchaisri, W., & Yamamotob, K. (2014). Biomass Production from Fermented Starch Wastewater in a Photo-bioreactor with Internal Overflow Recirculation. *Bioresource Technology*, *165*, 129-136. https://doi.org/10.1016/j.biortech.2014.03.119

Pradhan, A., Bhaumik, P., Das, S., Mishra, M., Khanam, S., Hoque, B. A., Mukherjee, I., Thakur, A. R., & Chaudhuri, S. R. (2008). Phytoplankton Diversity as an Indicator of Water Quality for Fish Cultivation. *American Journal of Environmental Sciences*, *4*(4), 406-411. https://doi.org/10.3844/ajessp.2008.406.411

Qin, L., Liu, Q., Meng, Q., Fan, Z., He, J., Liu, T., Shen, C., & Zhang, G. (2017). Anoxic Oscillating MBR for Photosynthetic Bacteria Harvesting and High Salinity Wastewater Treatment. *Bioresource Technology*, *224*, 69-77. https://doi.org/10.1016/j.biortech.2016.10.067

Talaiekhozani, A., & Rezania, S. (2017). Application of Photosynthetic Bacteria for Removal of Heavy Metals, Macropollutants and Dye from Wastewater: A Review. *Journal of Water Process Engineering*, *19*, 312-321. https://doi.org/10.1016/j.jwpe.2017.09.004

Thakur, R. K., Jindal, R., Singh, U. B., & Ahluwalia, A. S. (2013). Plankton Diversity and Water Quality Assessment of Three Freshwater Lakes of Mandi (Himachal Pradesh, India) with Special Reference to Planktonic Indicators. *Environmental Monitoring and Assessment*, *185*(10), 8355-8373. https://doi.org/10.1007/s10661-013-3178-3

Venkatesharaju, K., Ravikumar, P., Somashekar, R. K. & Prakash, K. L. (2010). Physico-chemical and Bacteriological Investigation on the river Cauvery of Kollegal Stretch in Karnataka. *Journal of Science Engineering and Technology*, *6*, 50-59.

Verlecar, X. N., & Desai, S. R. (2004). *Phytoplankton identification manual*. National Institute of Oceanography.

Walsh, G. E. (1978). Toxic effects of pollutants on plankton. In Butler, G.C. (Ed.). *Principles of Ecotoxicology* (Chapter 12; p. 257-274). Wiley.

Wu, J. T. (1984). Phytoplankton as a Bioindicator for Water Quality in Taipei. *Botanical Bulletin of Academia Sinica*, *25*(2), 205-214.

Yang, A., Zhang, G., Meng, F., Zhang, P., & Chen, Y. (2018). Membrane Concentrate Treatment by Photosynthetic Bacteria: Feasibility and Tolerance Mechanism Analysis. *Bioresource Technology*, *253*, 378-381. https://doi.org/10.1016/j.biortech.2018.01.034

Yusuf, Z. H. (2020). Phytoplankton as Bioindicators of Water Quality in Nasarawa Reservoir, Katsina State Nigeria. *Acta Limnologica Brasiliensia*, *32*, e4. https://doi.org/10.1590/S2179-975X3319

Zannatul, F., & Muktadir, A. K. M. (2009). A Review: Potentiality of Zooplankton as a Bioindicator. *American Journal of Applied Sciences*, *6*(10), 1815-1819. https://doi.org/10.3844/ajassp.2009.1815.1819

Zhou, Q., Zhang, P., & Zhang, G. (2014). Biomass and Carotenoid Production in Photosynthetic Bacteria Wastewater Treatment: Effects of Light Intensity. *Bioresource Technology*, *171*, 330-335. https://doi.org/10.1016/j.biortech.2014.08.08

Zhou, Q., Zhang, G., Lu, Y., & Wu, P. (2015). Feasibility Study and Process Optimization of Citric Acid Wastewater Treatment and Biomass Production by Photosynthetic Bacteria. *Desalination and Water Treatment*, *57*, 6261-6267. https://doi.org/10.1080/19443994.2015.1005687

Ziadi, B., Dhib, A., Turki, S., & Aleya, L. (2015). Factors Driving the Seasonal Distribution of Zooplankton in a Eutrophicated Mediterranean Lagoon. *Marine Pollution Bulletin*, *97*, 224-233. https://doi.org/10.1016/j.marpolbul.2015.06.012

RESEARCH ARTICLE
Journal of Biometry Studies (2024) 4(1): 11-30
DOI: 10.61326/jofbs.v4i1.02
www.biometrystudies.com

Journal of Biometry Studies
PRENSIP
www.prensip.gen.tr

Enhancing microbiological stability in municipal water distribution: A descriptive statistical analysis for public health assurance

Mostafa EISSA[1,*]

[1]Independent Researcher, Pharmaceutical & Public Health Research Facility, Biometry Studies Department, Cairo/EGYPT

[*]Corresponding author: mostafaessameissa@yahoo.com
Received: 24/01/2024, Accepted: 04/06/2024

Abstract

The microbiological quality of municipal water is of prime importance for public health and represents a prominent concern for authorities to maintain and control the bioburden level within safe limits. An appropriately functioning network must ensure the biological stability of the water distribution system throughout the way of water from the processing and purification station until reaching the ultimate consumers. The study examines methodologies used to evaluate different segments from a local distribution system of city water by examining short-term results of the total microbial count obtained from several available points of use through a one-year period. Microbial variations observed in city's water system. Autumn has the highest bioburden. Locations exceeding 10,000 CFU/100 mL threshold warrant investigation. Non-uniform distribution across the system suggests potential influence of distance, pipe type, and biofilms. Early warning system established to monitor bioburden surges. Descriptive analysis showed that six lines from 19 examined sections from the distribution system had lower dispersion in the microbiological data between minimum and maximum results than the remaining parts of the distribution pipes. Moreover, multiple comparison test outputs returned significant differences in microbiological density datasets between high-count and low-count use points. Thus, surveying the layout design of the municipal water distribution infrastructure would be crucial to restate its impact on the sanitary distribution of water. In conclusion, this refined study provides a nuanced examination of microbiological stability in municipal water distribution, offering targeted insights for effective public health management. By implementing the present assessment, public health authorities can spot and identify defective areas in the distribution system and take necessary actions to prevent or mitigate potential risks of waterborne diseases.

Keywords: Bioburden, Municipal water, Correlation matrix, Coefficient of variation, Pareto chart

Please cite this article as follows:
Eissa, M. (2024). Enhancing Microbiological Stability in Municipal Water Distribution: A Descriptive Statistical Analysis for Public Health Assurance. *Journal of Biometry Studies*, *4*(1), 11-30. https://doi.org/10.61326/jofbs.v4i1.02

1. Introduction

Water is essential for life, health, and well-being. However, not all water sources are equally safe and reliable. Municipal water, or tap water, is the main source of water for various industries and homes in many countries. It is delivered through underground pipes that connect the water treatment plants to the end-users. Municipal water undergoes rigorous treatment and processing to remove most of the impurities and contaminants that may pose health risks or affect the water quality. Municipal water is tap water that is delivered to various industries and homes via underground pipes (Sensorex, 2020). This water is fully treated and processed before being delivered to these destinations, which means that the majority of impurities are removed before it can be potable or used in the shower (Netsol Water Solutions Pvt, 2022). Even though the Environmental Protection Agency (EPA) has established water quality standards that each municipality must meet, the quality of tap water can vary from community to community (Netsol Water Solutions Pvt, 2022; US EPA, 2013). Therefore, it is important to monitor and evaluate the quality of municipal water and to identify and address the potential challenges and risks that may compromise its safety and suitability for different purposes. The goal of public health authorities is to assess the microbiological stability of municipal water distribution in urban regions, using various indicators and methods. The methodologies should provide valuable insights for effective public health management and water quality improvement (Apure, 2022).

Through a network of pipelines, reservoirs, and valves, municipal water distribution systems provide treated water to a variety of residences and businesses. This water's microbiological stability is crucial to guarantee its safety and purity for use and consumption by humans. The idea of giving customers drinking water with the same microbial quality at the tap as it produces at the water treatment facility is known as “microbiological stability” (Sensorex, 2020). However, during distribution, unchecked bacterial and other microbe development may happen, posing a risk to hygienic conditions as well as aesthetic and functional issues (Netsol Water Solutions Pvt, 2022). Microbial proliferation, for instance, can lead to the emergence of opportunistic diseases, like Legionella, which can endanger the health of susceptible people (US EPA, 2013). Additionally, microbial development can change the water's color, taste, odor, and turbidity.

In general, drinking municipal water must be relatively safe. For a municipality to provide water to residents and businesses, the water must first be treated and tested (CDC, 2022). If it does not meet local or national standards, the water must be treated again until a certain number of contaminants have been removed (World Health Organization, 2008). As previously stated, some municipalities are more effective than others at treating tap water, which means that certainty about the quality of municipal water cannot be assured from the final consumer point of view (CDC, 2020). The concept of biological stability of drinking water refers to providing consumers with drinking water of the same microbial quality as that produced at the water treatment facility (Park et al., 2016; Prest et al., 2016). Uncontrolled growth of bacteria, on the other hand, can occur during distribution in water mains and premise plumbing, resulting in hygienic (e.g., development of opportunistic pathogens), aesthetic (e.g., deterioration of taste, odor, color), or operational (e.g., fouling or bio-corrosion of pipes) issues (Hayward et al., 2022; Ibekwe & Murinda, 2019). Diverse microorganisms compete for limited available nutrients for growth in drinking water (Hibbing et al., 2010).

Water is an essential resource, and ensuring its safety is a top priority for municipalities around the world (Eissa, 2017). One critical aspect of this is assessing microbiological stability, which involves analyzing the presence of microorganisms in water distribution systems (Cabral, 2010). Descriptive statistical analysis plays a crucial role in this process, as it helps to identify trends and patterns in the data (Eissa, 2015). By using statistical tools such as mean, standard deviation, and correlation analysis, experts can gain valuable insights into the health of these systems and make informed decisions about how to maintain and improve their performance (Eissa, 2016a). In this article, we will explore the importance of descriptive statistical analysis in assessing microbiological stability in municipal water distribution systems, and how this approach can help to ensure the safety and well-being of communities around the world. The importance of descriptive statistical analysis in assessing microbiological stability in municipal water distribution systems will be explored in this article, and how this approach can help to ensure the safety and well-being of communities around the world. However, it is also acknowledged that the validity and reliability of this method depend on the sample size and frequency of testing, which may be constrained by practical and financial factors in the initial phases of surveillance.

The safety and quality of drinking water are of utmost importance, and as such, it is crucial to ensure that municipal water distribution systems are microbiologically stable (Eissa, 2018). One way to assess this is through the current in-depth analysis, which allows us to better understand the trends and patterns in microbial data. By examining key factors such as microbial load and variation in the count, potential sources of defects should be identified and proactive measures to prevent them.

The microbial quality of water is critical for the validity of human use, especially since humanity is approaching an advanced stage of water scarcity. Due to the challenges encountered with city water use and its safety microbiologically, an effective examination technique, including quantitative and descriptive methodologies must be adopted by the public health authority to assess the microbial quality of the municipal distribution system. A focus on a specific topic concerning biological stability shall be discussed herein via bioburden density in tap water to measure the magnitude of the stability of this inspection property at the preliminary stages of monitoring with a limited number of samples available. This study aims to assess the microbiological stability of municipal water distribution systems in a selected city, using descriptive statistical analysis of microbial data. This approach allows us to better understand the trends and patterns in microbial data and to identify key factors that influence the microbial load and variation in the water. By examining these factors, we can detect potential sources of defects and propose proactive measures to prevent them.

2. Material and Methods

Municipal water samples were collected monthly for one year when available or accessible.

2.1. Location and study subject

A specific area embracing a tape water distribution network was selected for investigation of the total microbial aerobic count. The examined piping system is contained within a single district included within the Giza governorate. The study included 19 sampling points of use after two storage tank units from the processed city water. The selection of sampling points and criteria considered several factors. First, the availability and accessibility of data on the water distribution network and water quality parameters within the district were crucial. Additionally, the representativeness and diversity of water consumption patterns and water demand allocation methods within the district were considered. The research objectives and proposed model also played a role, with the relevance and applicability of the chosen district being a key factor.

Spatial distribution and coverage of sampling points were crucial. Points were chosen to represent different zones and branches of the network, ensuring a comprehensive overview. Finally, the functional and operational characteristics of the points of use themselves were taken into account. This included factors like the type of water use (domestic, commercial, etc.), water consumption frequency and duration, and the presence of storage tanks or other devices that could influence water quality. Only points with reliable water quality measurements, such as total microbial aerobic count and turbidity, were included.

2.2. Type of microbiological analysis

The microbiological quality of city water samples was assessed by measuring the bioburden density in 100 mL of water as Colony Forming Unit (CFU). The samples were collected monthly from the available points of use throughout the year 2022. The Total Viable Aerobic Count (TVAC) was determined by filtering the samples through 0.45 μm sterile membrane filters and incubating them on appropriate culture media (Eissa et al., 2022). The neutralization was performed by adding sodium thiosulfate to the sample containers, as recommended by the ISO 19458:2006 standard on the sampling bottles (Eissa & Rashed, 2022; Eissa et al., 2022).

2.3. Statistical analysis and software platforms

A variety of statistical techniques were employed to analyze the microbiological data, including Pareto analysis, outlier detection, and non-parametric tests like the Wilcoxon signed-rank test (for comparing medians of paired samples) and the Kruskal-Wallis test (for comparing medians of multiple groups). The Spearman correlation matrix was also used to assess relationships between microbial levels at different use points. These techniques helped identify significant differences in bioburden levels across seasons, locations within the water network, and even specific use points.

3. Results

3.1. Water microbial density and time factor

The analysis of microbial counts in the city's water system revealed significant seasonal variations. Figure 1 utilizes the Pareto principle to highlight the most impactful seasons and months on the overall bioburden. When combined, spring and autumn bioburden levels contribute over 80% of the total microbial count. Further breakdown by month shows an even greater disparity. Two fall months, September and November, account for more than half (54%) of the total bioburden, while two spring months, April and May, contribute only 23.5%. Interestingly, even summer's August month falls within the top contributors, accounting for an additional 7.5%, bringing the combined impact of these three seasons to over 85% of the total microbial count.

3.2. Microbiological variations and extremities over the study term for the analyzed point-of-use

The procedure of sample preparation, collection and processing was consistent with previous similar research that was conducted on the same line (Eissa & Rashed, 2022). This procedure was chosen because it is a standard and reliable method for detecting and enumerating the total microbial population in water samples (Eissa & Rashed, 2022; Eissa et al., 2022). It is also important to neutralize the residual disinfectant in water samples before analysis, as it can inhibit the growth of microorganisms and affect the accuracy of the results. Aberrant Samples Detection for the water use ports labeled SK1, ST2, RNT, LER, PCW, MIQ, G02, G04, G07, G3A, G3B, G18, G19, G33, G52, G68, G71, G75 were statistically tested as could be seen in Figure 2. The number of outliers detected was 1, 1, 1, 2, 2, 2, 2, 2, 3, 1, 2, 2, 2, 3, 2, 1, 0, 0, 1, respectively. Thus, the excursion rates (%) were 8.3%, 9.1%, 8.3%, 16.7%, 18.2%, 16.7%, 16.7%, 25.0%, 8.3%, 16.7%, 16.7%, 16.7%, 25.0%, 16.7%, 8.3%, 0.0%, 0.0%, 8.3%, respectively. This finding was summarized in Figure 2 with two extreme rates (maximum 25.0% and minimum 00.0%) of the excursion rates shown in addition to the statistics computed in the clustered column diagrams.

The Coefficient of Variation (CV) is a statistical measure that relates the standard deviation of a set of values to its mean. It is often expressed as a percentage and is used herein to compare the degree of variability between distinct sets of microbiological counts in a water line (Figure 3). For estimated values, the CV can be used to assess the precision of the. A lower CV indicates a more precise estimation of water quality (G68, G07, G71 and G03), while a higher CV suggests greater uncertainty or variability around the estimate (ST2, G18, G02 and PCW). By expressing the coefficient of variation as a percentage, it provides a standardized measure of dispersion that can be used to compare data sets across different domains. A fluctuation in the coefficient of variation is a useful statistical tool that can help investigators and analysts to assess the degree of variability and precision in the water distribution system datasets.

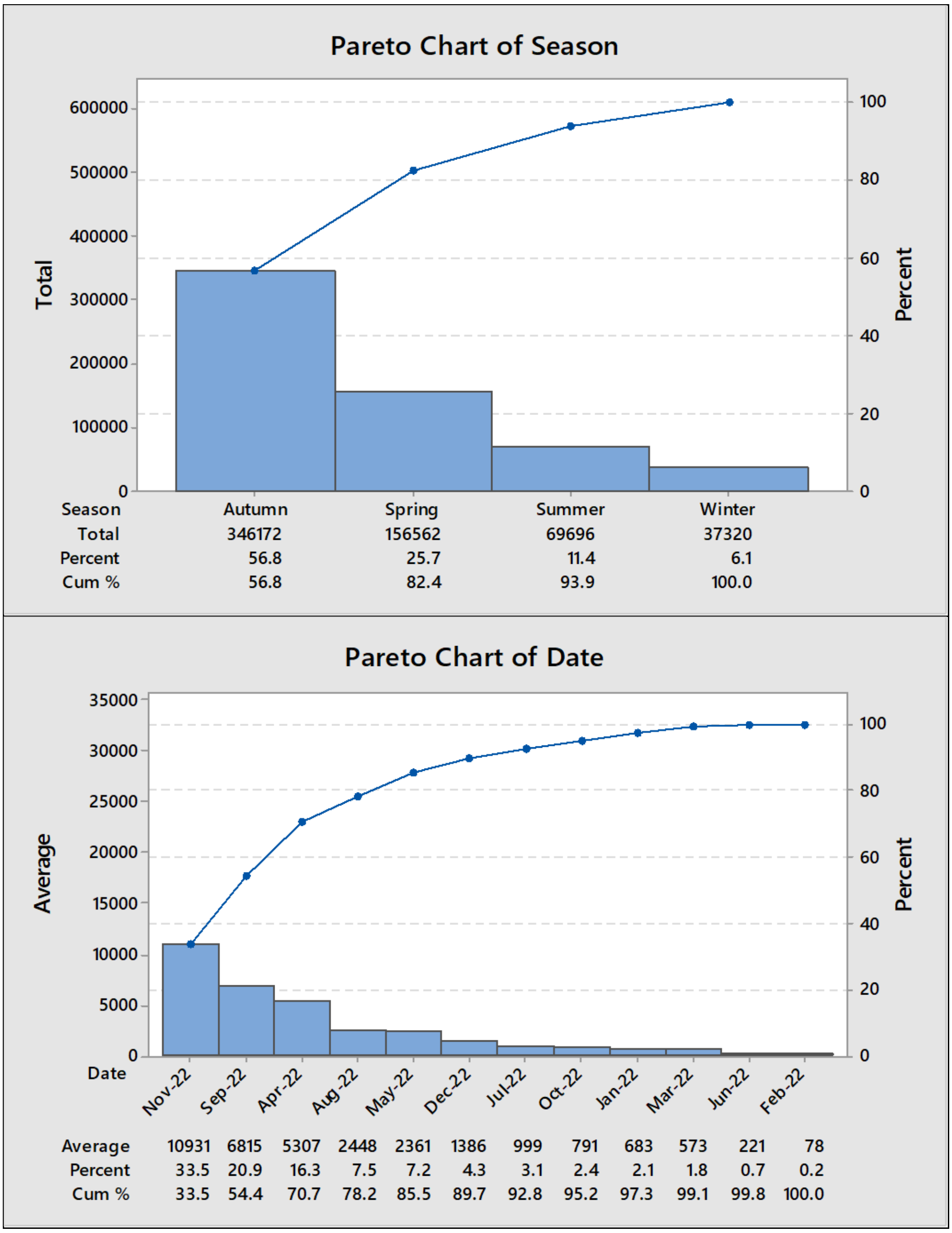

Figure 1. Pareto analysis showing the periods embracing major bioburden level in city water distribution system

The first diagram presents two Pareto charts that illustrate the bioburden levels in a city's water distribution system across various seasons and dates. Bioburden refers to the quantity of microorganisms present in a sample. Figure 1 depict the total or average bioburden levels for each category, along with the cumulative percentage of the total bioburden. The categories are organized in descending order of bioburden levels, from left to right. These charts are useful in identifying the most significant factors that contribute to the bioburden levels in the water system. The upper chart displays the total bioburden levels for each season. Autumn records the highest bioburden level, followed by spring, summer, and winter. This suggests that the bioburden levels fluctuate with the seasonal changes in temperature, rainfall, and water demand. The lower chart reveals the average bioburden levels for specific months. It is observed that some dates have significantly higher bioburden levels than others, which could be attributed to factors such as water quality, maintenance, or contamination events.

The second diagram demonstrates the results of outlier detection of microbial count in the examined distribution lines of municipal water. The upper graph uses vertical bars to represent microbial counts at different use points labeled from SK1 to G75. The blue bars indicate outliers, while the red bars indicate non-outliers. The y-axis is labeled as "Frequency" ranging from 0 to 12. The lower graph is another bar plot representing the distribution of microbial counts at different use points. Figure 2 shows the interquartile range (IQR), lower limits, and upper limits using blue, red, and green lines respectively. The y-axis is also labeled as "Statistical Limits" but has a different scale x 10,000. This diagram illustrates that the microbial count varies significantly across the use points, and that some use points have extremely high or low values that deviate from the normal range.

Figure 3 depicts the Coefficient of Variation (CV) for bioburden samples estimated from different locations from the distribution lines of city water under examination. The CV values are represented on a bar graph, with each bar corresponding to a specific location or sample type. The CV is a measure of the relative variability of a data set, expressed as a percentage. It is calculated by dividing the standard deviation by the mean and multiplying by 100. A higher CV indicates a higher degree of dispersion or heterogeneity in the data, while a lower CV indicates a higher degree of consistency or homogeneity in the data. The graph shows that the CV values vary widely across the different locations and sample types. This suggests that the bioburden levels in the city water are not uniform and may depend on factors such as the distance from the source, the type of pipe, the presence of biofilms, and the sampling method.

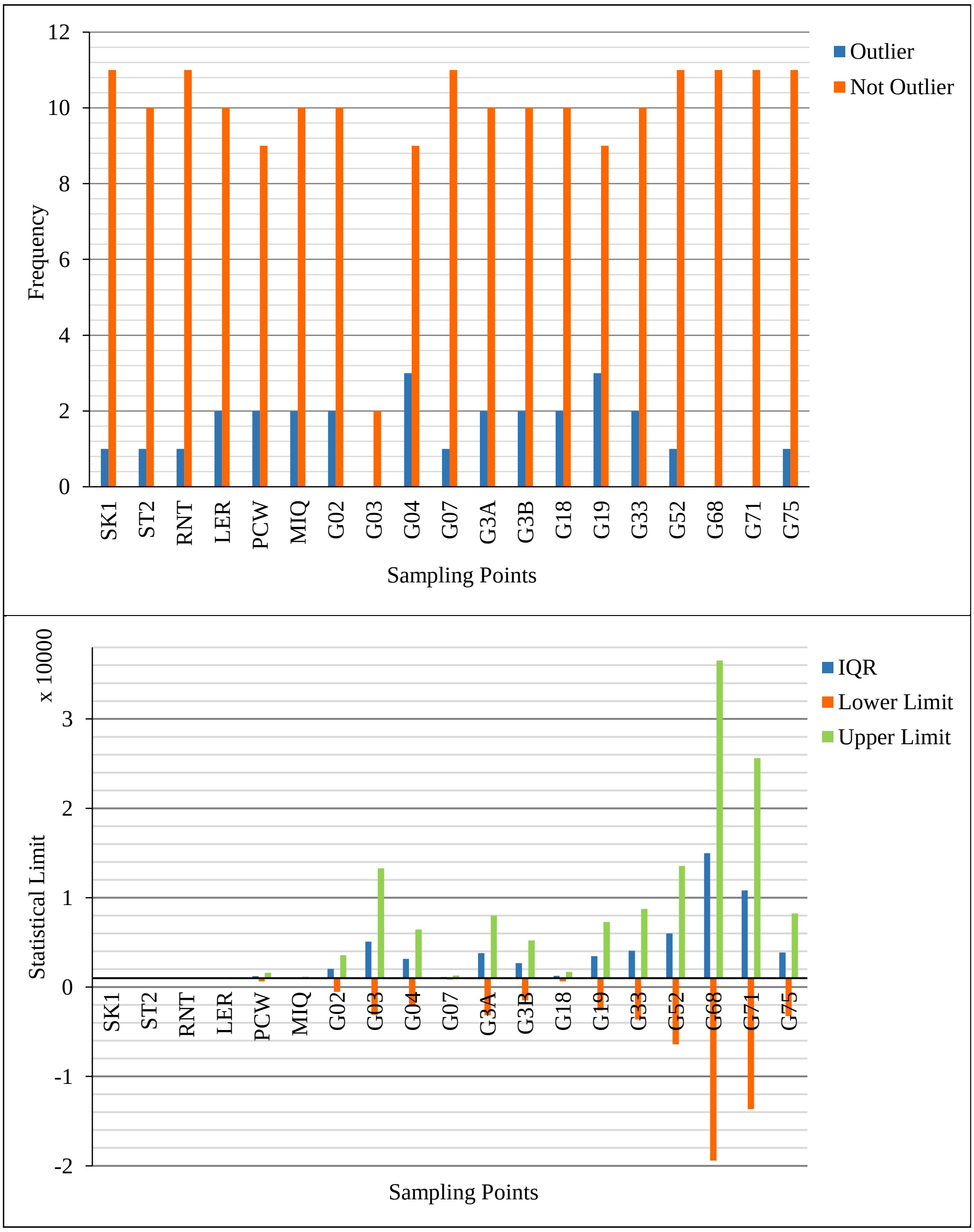

Figure 2. Outlier detection of microbial count in the examined distribution lines of municipal water

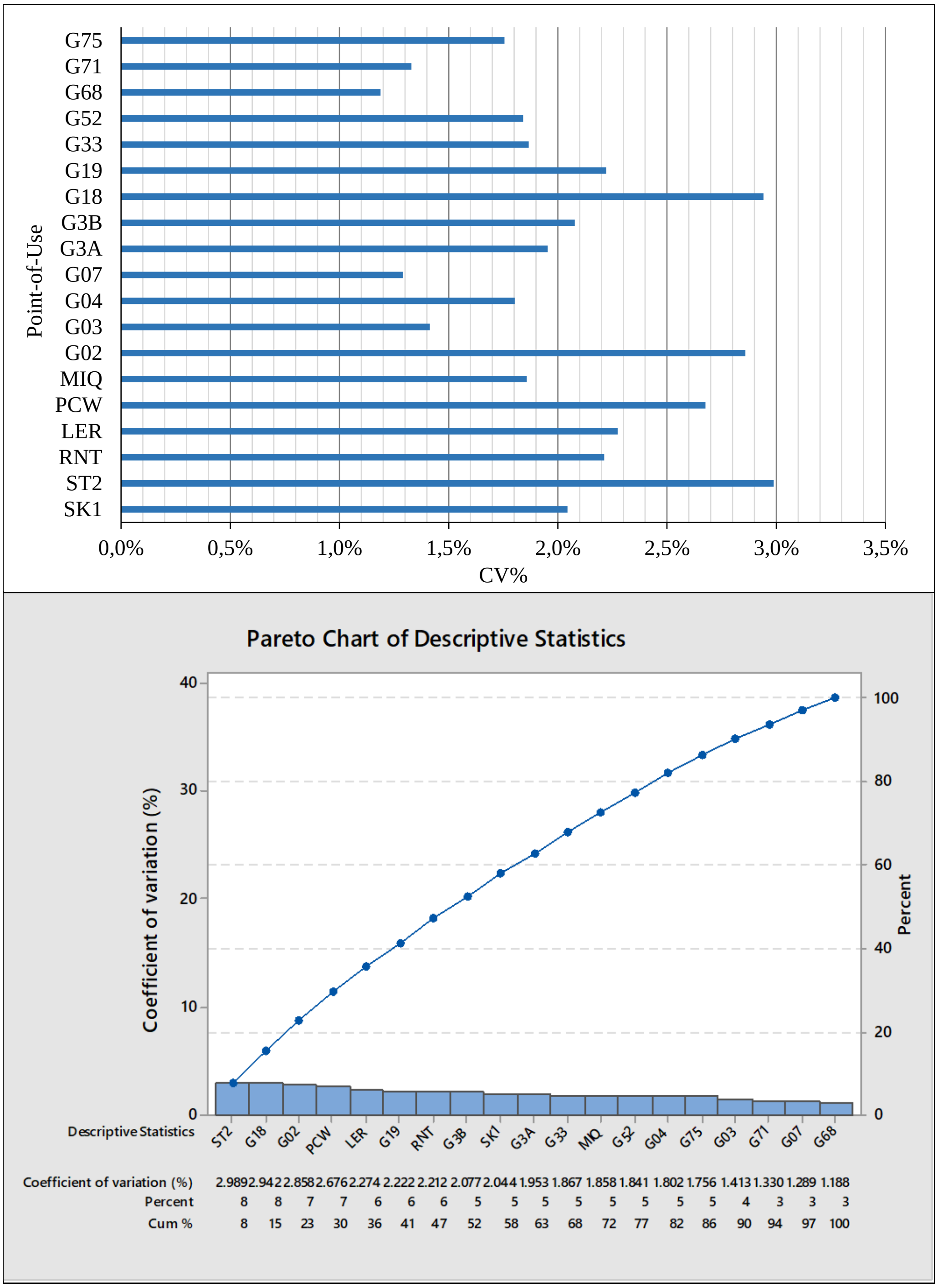

Figure 3. Coefficient of variation (CV) for the bioburden samples estimated from different locations from the distribution lines of city water under examination

3.3. Establishment of early warning alarm using cut-off threshold

All Setting a warning limit for the microbiological count of city water distribution lines would be a desirable measure. This cut-off limit was set arbitrarily far below the action threshold of 50,000 CFU/100 mL value namely 10,000 CFU/100 mL. The non-parametric *t*-test analog test was used to identify the lines with a higher potential of bioburden density. With this respect distribution segments that showed actual medians that are statistically not significantly different from the reference value at α=0.05 were the main source of the microbial density aberrancy i.e., use ports G68, G71 and G3A with minority in G19 and G33. This is reflected in the discrepancy values and indicated by the two-tailed *p*-value as could be seen in Figure 4. The remaining points of use were far below this alarming limit and hence are statistically different from it. This study established an early warning alarm for the microbiological count of city water distribution lines using a cut-off threshold of 10,000 CFU/100 mL. This threshold was chosen to be much lower than the action threshold of

50,000 CFU/100 mL to ensure timely detection and prevention of microbial contamination. By applying a non-parametric test, the study identified the distribution segments that had a higher potential of bioburden density and were not significantly different from the reference value. These segments were G68, G71, and G3A, with some contribution from G19 and G33, as shown in Figure 4. These segments should be monitored closely and subjected to appropriate interventions to reduce the bioburden levels and improve the water quality. The remaining points of use were well below the alarming limit and were significantly different from the reference value, indicating a satisfactory microbial stability.

Figure 4 shows the results of a non-parametric *t*-test analog that compares the median microbial counts of different water lines with a reference value of 10,000 CFU/100 mL, which is the cut-off threshold for the early warning alarm. The *p*-values indicate the probability of obtaining a result equal to or more extreme than the observed one, assuming that the null hypothesis is true. The null hypothesis is that the median microbial count of each water line is equal to the reference value. The lower the *p*-value, the more evidence there is to reject the null hypothesis and conclude that the water line has a significantly different median microbial count from the reference value. The figure illustrates that some water lines have major or minor microbial counts that are statistically referenced to the alarming threshold value, meaning that they have significantly higher or lower median microbial counts than the reference value. These water lines are G68, G71, G3A, G19, and G33, as shown by their low *p*-values and high discrepancies. These water lines should be monitored closely and subjected to appropriate interventions to reduce the bioburden levels and improve the water quality. The remaining water lines have median microbial counts that are not significantly different from the reference value, indicating a satisfactory microbial stability.

3.4. Descriptive statistical distribution of microbiological data

The descriptive statistical analysis of the microbial count per 100 mL of water for different sampling locations is shown in Figure 5 in descending order of magnitude. The datasets demonstrated non-Gaussian distribution with strong skewness to the right side. The significance of demonstrating a non-Gaussian distribution for microbial count data is that it indicates that the data are not symmetric and have a long tail on the right side, meaning that there are some extreme values or outliers that are much higher than the average. This implies that the data are over-dispersed and have more variability than expected from a normal distribution. This affects the interpretation of the data because it means that the mean and standard deviation are not sufficient to describe the central tendency and dispersion of the data and those parametric statistical tests that assume normality may not be valid or reliable. Therefore, alternative methods such as non-parametric tests, transformation, or modeling with appropriate distributions (such as Poisson, negative binomial, or zero-inflated models) may be needed to analyze the data and draw meaningful conclusions.

This distribution is important in the context of water quality assessment because it reflects the heterogeneity and complexity of the microbial communities in different water sources and environments. Microbial count data are influenced by many factors, such as sampling methods, sequencing technologies, environmental conditions, and biological interactions. These factors can cause variation in the abundance and diversity of microbes across samples and locations, resulting in skewed and over-dispersed data. Understanding the distribution of the data can help to identify the sources and patterns of microbial contamination, assess the risks and impacts of waterborne diseases, and evaluate the effectiveness of water treatment and management strategies Thus, most of the sampling ports showed results that did not pass the normality test (at a significance level α 0.05). The stability of the bioburden level could be demonstrated by points RNT, SK1, ST2, MIQ, LER and G07. However, rising level of variability could be explored in G03, PCW, G75, G04, G18, G3B, G71, G52, G68, G33, G3A, G02 and G19, respectively. The rising level of variability in the sampling ports indicates that the microbiological counts are not consistent and may fluctuate over time or space. Some factors that may contribute to this variability are:

- *Water quality and treatment*: The source water quality, the type and dosage of disinfectants and the effectiveness of filtration and chlorination may affect the microbial diversity and abundance in the water supply systems.
- *Water distribution system*: The age, material, and design of the pipes, the presence of biofilms, the water pressure and flow rate, and the occurrence of leaks or breaks may influence the microbial community composition and dynamics in the water distribution system.
- *Building water supply system*: The configuration, maintenance, and usage of the building water supply system, such as the faucets, showers, heaters, and coolers, may impact the microbial growth and survival in the building water supply system.
- *Environmental factors*: The temperature, seasonality, light irradiation, pH, nutrients, organics, and antibiotics may affect the microbial activity and diversity in the water supply system.

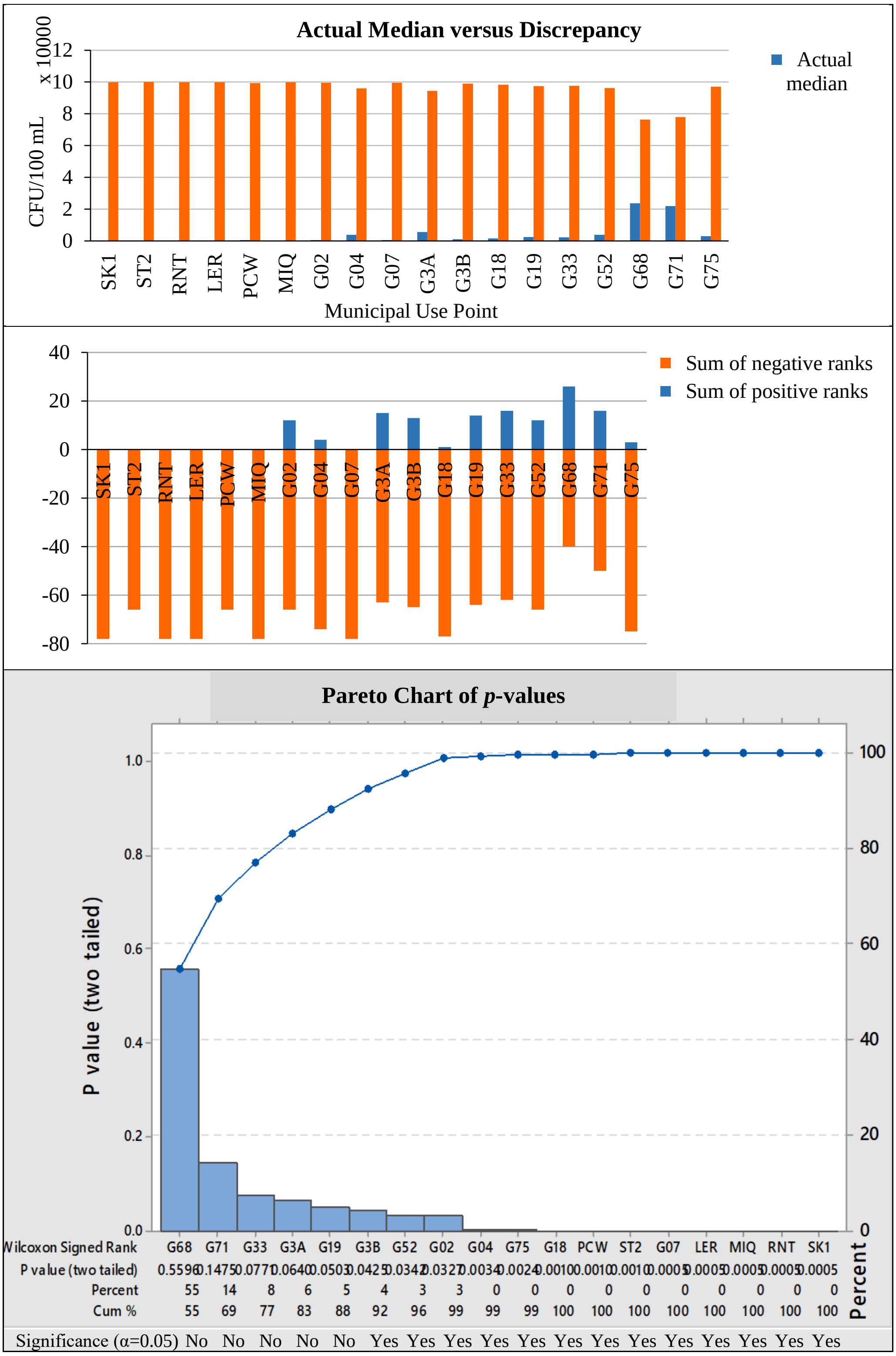

Figure 4. The output of the non-parametric *t*-test analog shows lines with major and minor microbial counts that are statistically referenced to the alarming threshold value. The top part of the figure displays a bar chart comparing the actual median and discrepancy at various municipal use points. The bars represent different values, with red indicating discrepancies and blue representing the actual median. The discrepancy is the difference between the actual median and the reference value. The middle part shows the sum of negative ranks for each municipal use point. The negative ranks are the ranks of the observations that are below the reference value. The bottom part, labeled as “Pareto Chart of *p*-Values”, illustrates the *p*-values associated with each municipal use point in a descending order.

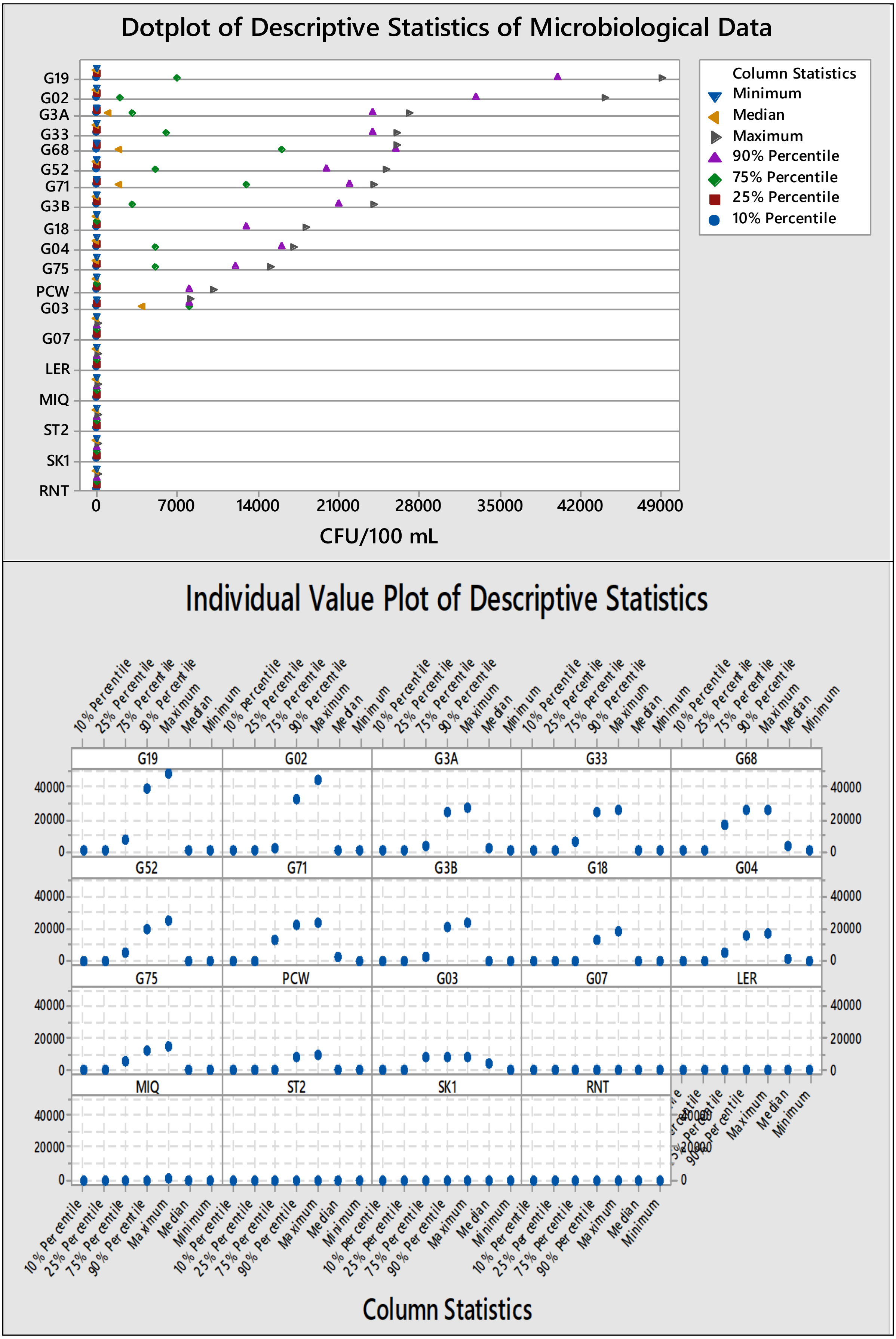

Figure 5. Column statistics showing data spreading of microbiological count from city water from different point-of-use

The two ways for data visualization mentioned in Figure 5 could be described as the following:

- *Dot plot of descriptive statistics*: This plot provides a visual representation of the spread and central tendency of microbiological data from different points-of-use. Each point-of-use is represented by different colored dots that indicate various percentiles, median, and minimum/maximum values. This visualization aids in quickly identifying the distribution and variability of microbial counts across different water lines, helping to assess their microbial stability.
- *Individual value plot*: This plot displays individual data points for each point-of-use, offering a detailed view of the microbiological count at each specific location. It allows for an in-depth analysis to identify any outliers or specific points with unusually high or low microbial counts, contributing to a comprehensive understanding of the water lines' microbial stability.

3.5. Pairwise comparison between microbiological record data from points of use under investigation

A water sample from one site, specifically the coast near UPV hatchery, was taken to the Microbiology Laboratory of IA-UPV for the total plate count (TPC) of bacteria present. Only one site was processed instead of two due to time constraints and the availability of materials in the laboratory. Five-fold dilutions were prepared with sterile saline solution. An aliquot (100 μL) of each sample dilution was spread onto plates with prepared media using an L-rod. The plates were then inverted, placed inside transparent plastic cellophane, and stored in the laboratory at room temperature for 72 hours. Only the first, third, and fifth dilutions were used in this study. Two replicates per sample were prepared and executed. All plates were monitored every 24 hours, and the number of colonies present was counted. For the calculation of bacteria colonies using the TPC method, the following procedure was employed.

Table 1 shows the main finding of the statistical test highlighting the presence of detectable differences in the bioburden quality between different lines of the distribution system. This finding is in line with the previous outcome of the column statistics that described the nature of data dispersion. Detailed examination of the sources of significant variability could be projected from Figures 6 and 7 which drew the mean rank difference for significant and non-significant pairwise comparisons, respectively.

Figure 6 shows the mean rank difference for various pairs of municipal water distribution segments, indicating the relative performance or condition of these segments in terms of microbiological density. Significant differences are marked with "Yes" in the Significance column and indicate that the segments have different levels of microbial contamination. Non-significant differences are not marked and indicate that the segments have similar levels of microbial contamination, with no clear advantage for either segment. Figure 7 shows the individual value plot of the mean difference in rank for the microbiological count of different water distribution system segments. The x-axis represents the Dunn's multiple comparisons test, which is a statistical method to compare the means of multiple groups after a non-parametric analysis of variance (ANOVA). The y-axis represents the mean difference in rank, which is the difference between the average ranks of two groups in a non-parametric test. The blue dots represent the individual value points of the mean difference in rank for each pair of water distribution system segments. The horizontal line at $y=0$ indicates no difference in rank between the two groups being compared.

The dot plot reveals variations in microbiological density across different segments of a municipal water distribution system. Some segments have consistently higher ranks (worse conditions) than other segments. This suggests that these segments may have different sources, pathways, or factors affecting their microbial quality. A few use points were the reason behind this significance since they contributed to unusually high microbial density in water samples. This could be illustrated in Figure 8 where the count of the main effectors in the municipal distribution system is small in relation to the other stable and low bioburden water lines. With reference to Figure 5, the points of discrepancy could be identified between the two distinct groups. Figure 9 shows two important criteria:

- The ranking of the microbiological sampling output results for each distribution port demonstrates the range, spreading and scatter of the data points.
- Histogram showing the similarity in the microbial count distribution shape and spreading that could be observed with the most points.

Table 1. The output of non-parametric multiple comparisons Kruskal-Wallis test for municipal water distribution system

Table Analyzed[¥]	Results of the distribution line data
Kruskal-Wallis test	
p-value	< 0.0001
Exact or approximate p-value?	Approximate[€]
p-value summary	****
Do the medians vary signif. (p<0.05)	Yes[§]
Number of groups	19
Kruskal-Wallis statistic	80.72
Data summary	
Number of treatments (columns)	19
Number of values (total)	214

[¥]Reported at 95% Confidence Interval (CI)
[€]Asymptotic p-value
[§]If the p-value is small, the hypothesis that the difference is due to random sampling can be rejected, and one can instead conclude that the populations have different distributions.
****A tiny p-value where $p \leq 0.0001$

Histogram of Bioburden: The upper graph shows the frequency of microbial counts across different distribution ports. Each bar represents the number of samples that fall within a certain range of microbial counts. The histogram helps to visualize the distribution and variation of microbial counts across different ports, and to compare the relative levels of contamination among them. Kruskal-Wallis Test Ranks: The lower graph shows the ranks of the microbiological data for each examined use point, based on the Kruskal-Wallis test. This is a non-parametric statistical test that compares the medians of multiple groups of data. The test assigns a rank to each data point, from the lowest to the highest value, and then calculates the average rank for each group. The lower the average rank, the lower the median of the group. The graph plots the average ranks for each distribution port, using different symbols. The graph helps to identify which ports have significantly higher or lower microbial counts than others, and to test the null hypothesis that all ports have the same median.

3.6. Non-parametric correlation matrix for the points of use of municipal water system

The apparent shape of the distribution in the histograms of Figure 8 showed truncated spreading of the microbial count data. Generally, the datasets do not follow a normal distribution. Hence, surveying the correlation between use points under examination was conducted using a non-parametric Spearman matrix. Table 2 shows the detailed output of the calculated two-tailed correlation coefficient (r) and the corresponding *p*-values with the strength of the association indicated and marked with visual color for ease of scanning and spotting of the significant correlation. The lowest negative but strong correlation was found between G07 and G3B. The correlation data of these points appeared to be highly associated. Section G3B appeared to be highly determined by G02 and PCW. On the other hand, SK1 and ST2 fields appeared to be highly correlated. Similarly, G02 with G33, MIQ with G07, G52 with G68, RNT with G18, ST2 with G52, MIQ with G3B and G19, G52 and G18 with LER and G68 with G71 correlation data were found to be highly linked together. A very strong positive correlation existed between G52 and ST2 in the bioburden data and was sufficiently strong with SK1. A strong correlation appeared between G71 and G68 but they do not share this strength level with other points.

4. Discussion

Water is one of the most essential resources on the planet, and its quality is crucial for human health and well-being (Bruyninckx, 2018). Municipal water distribution systems play a vital role in ensuring that people have access to safe and clean drinking water (Dinka, 2018). However, these systems can be vulnerable to microbiological contamination, which can pose serious health risks to the public (Vacs Renwick et al., 2019). That's why it's crucial to conduct regular assessments of the microbiological stability of water distribution systems (Liu et al., 2018). Descriptive statistical analysis is an important tool that can help experts to interpret and analyze the data collected from these assessments (Noiva et al., 2016). In the following sections, we'll explore the importance of descriptive statistical analysis in assessing the microbiological stability of municipal water distribution systems and how it can be used to identify potential risks and improve the quality of drinking water. The biological stability of municipal water is of prime importance for public health use (Pruden, 2014). Thus, safety must be ensured for human use (WHO, 2011). One of the important criteria that should be monitored and evaluated is the microbiological content of the water. The bioburden level should be ensured to be within the permissible limit from the departure from the station until reaching the consumers (Health Canada, 2021). The microbiological efficacy of the distribution system could be examined through the magnitude of the bioburden density and the stability of the microbial enumeration over a specific time and from one segment to another.

4.1. Water microbial density and time factor

The provided information makes a compelling case for the seasonal variation in microbial count and its implications for public health. The use of the Pareto principle in Figure 1 highlights the significant impact of certain seasons and months on bioburden, indicating the need for targeted interventions. The reference to Edam and Abdelgalil (2022) underscores the empirical basis of the findings, adding credibility to the assertion that autumn contributes significantly to the total bioburden. Moreover, the combination of autumn and spring bioburden accounting for over 80% of the total microbial count emphasizes the concentrated impact of specific seasons on water quality.

Edam and Abdelgalil (2022) is a recent study that investigated the seasonal variation of microbial community, potential opportunistic pathogens and antibiotic resistance genes in a novel underground reservoir of an island. The study found that the microbial community and function showed significant seasonal changes, with autumn having the highest diversity and abundance of microbes. The article summarized the current knowledge on the microbial ecology of water and its relation to public health. The researchers discussed the factors that influence the microbial composition and dynamics of water, such as temperature, nutrient availability, hydrological conditions, and human activities. The paper also highlighted the potential risks of waterborne pathogens and the need for effective monitoring and management of water quality.

The detailed breakdown by month further strengthens the argument, demonstrating the disproportionate contribution of certain months to the overall bioburden. The reference to Figueras and Borrego (2010) supports the significance of seasonal impact on water quality, emphasizing the need for tailored sanitization and disinfection strategies to manage bioburden levels effectively. In conclusion, the evidence presented suggests that seasonal variations have a substantial impact on the biological quality of city water, necessitating careful consideration and targeted action by public health authorities to maintain control over water quality and safeguard public health.

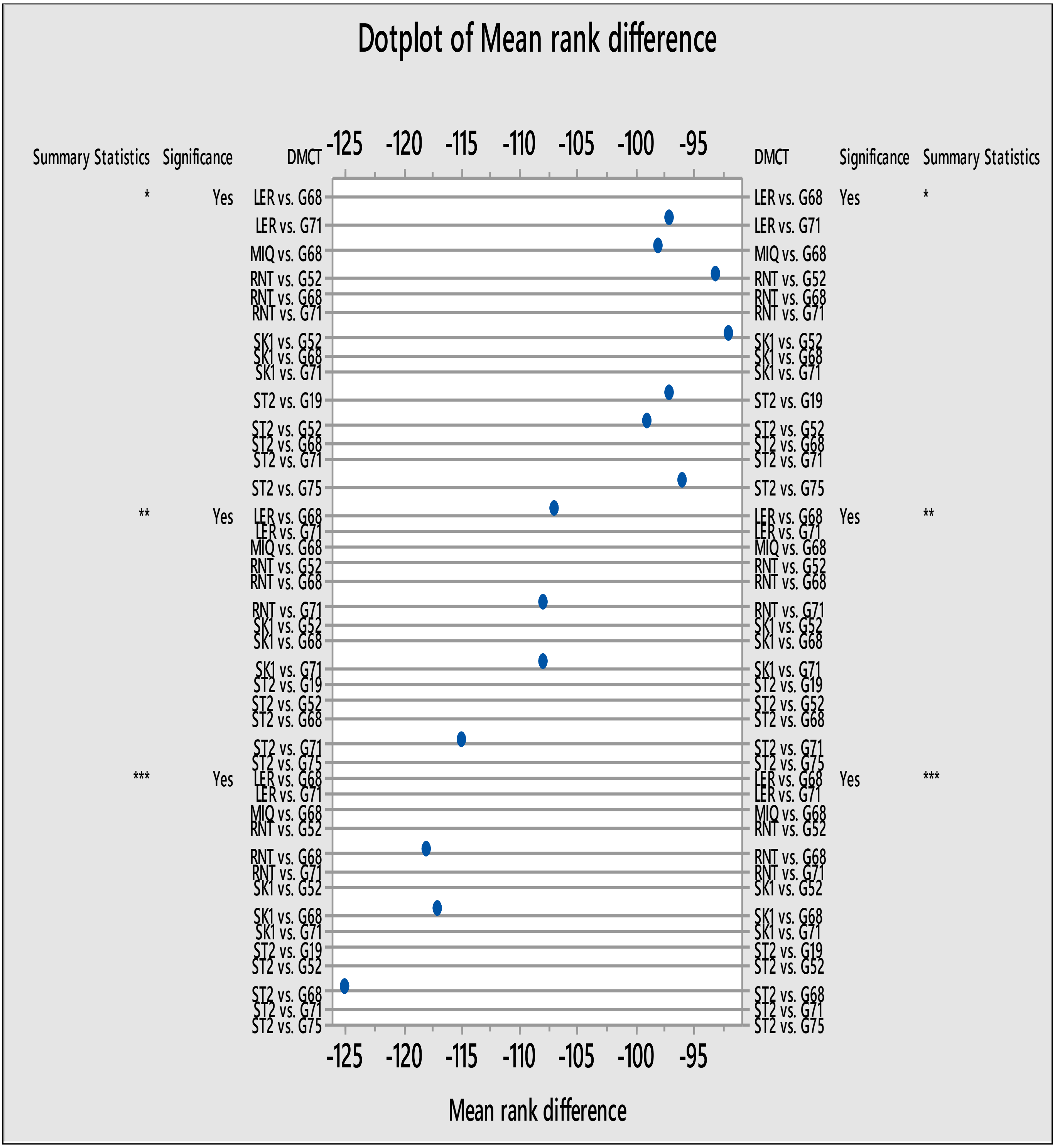

Figure 6. The outcome of the mean rank difference from the statistically significantly different segments of the water distribution system in the microbiological density

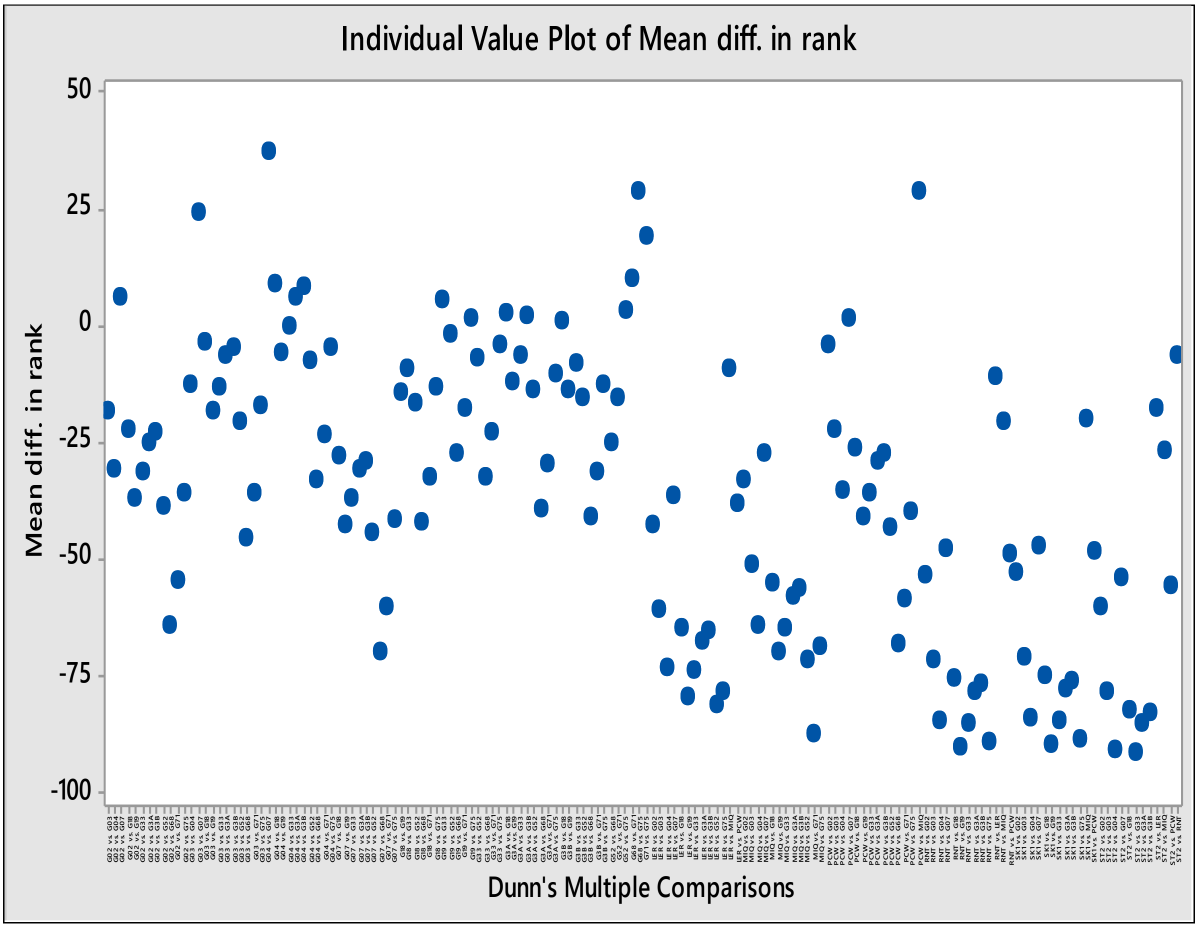

Figure 7. The outcome of the mean rank difference computed for the statistically not different water distribution system segments in the microbiological count

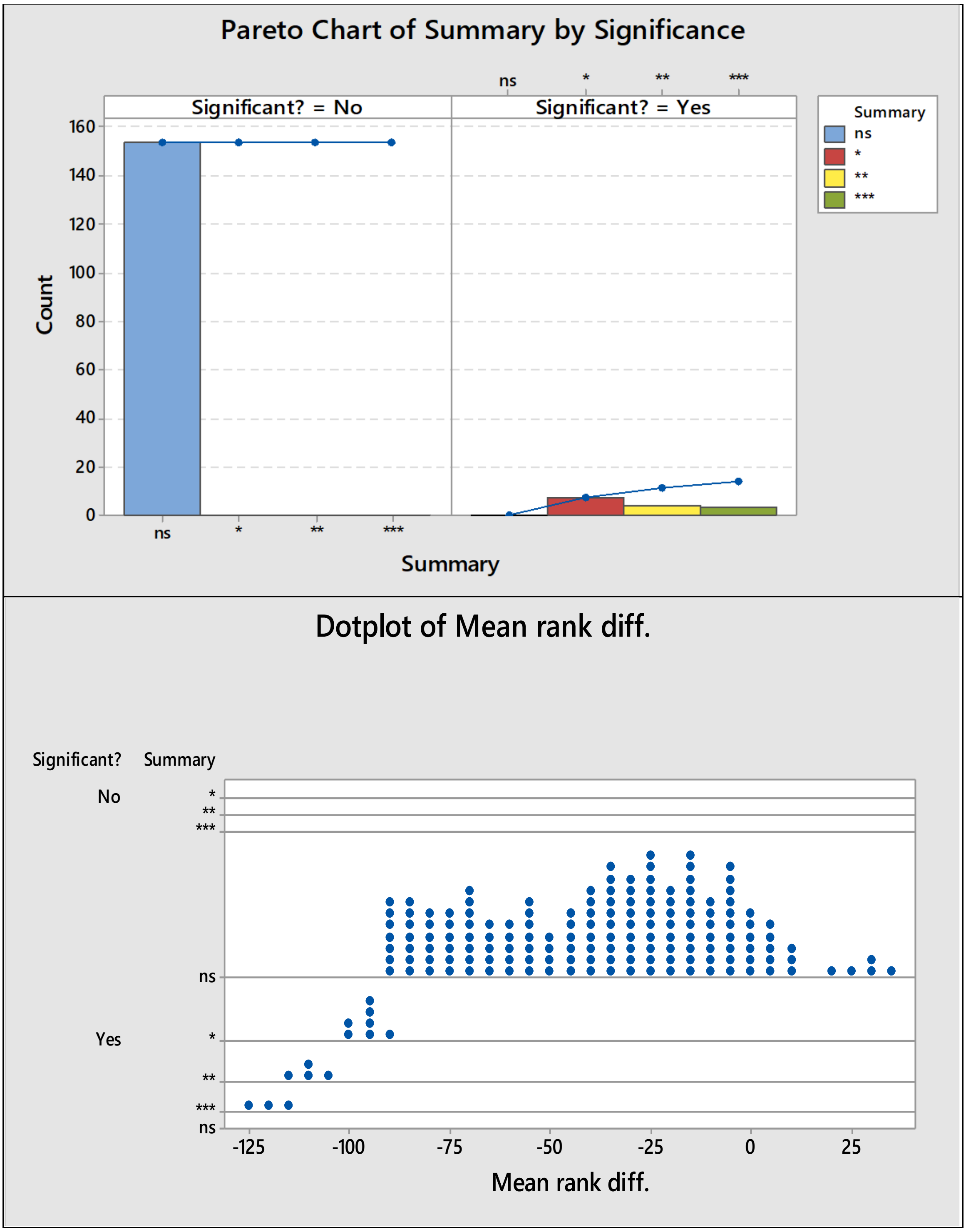

Figure 8. Mean rank analysis of the non-parametric Kruskal-Wallis test summing up the multiple comparison results (ns: not significant (p>0.05), *p≤0.05, ** p≤0.01) and ***p ≤0.001)

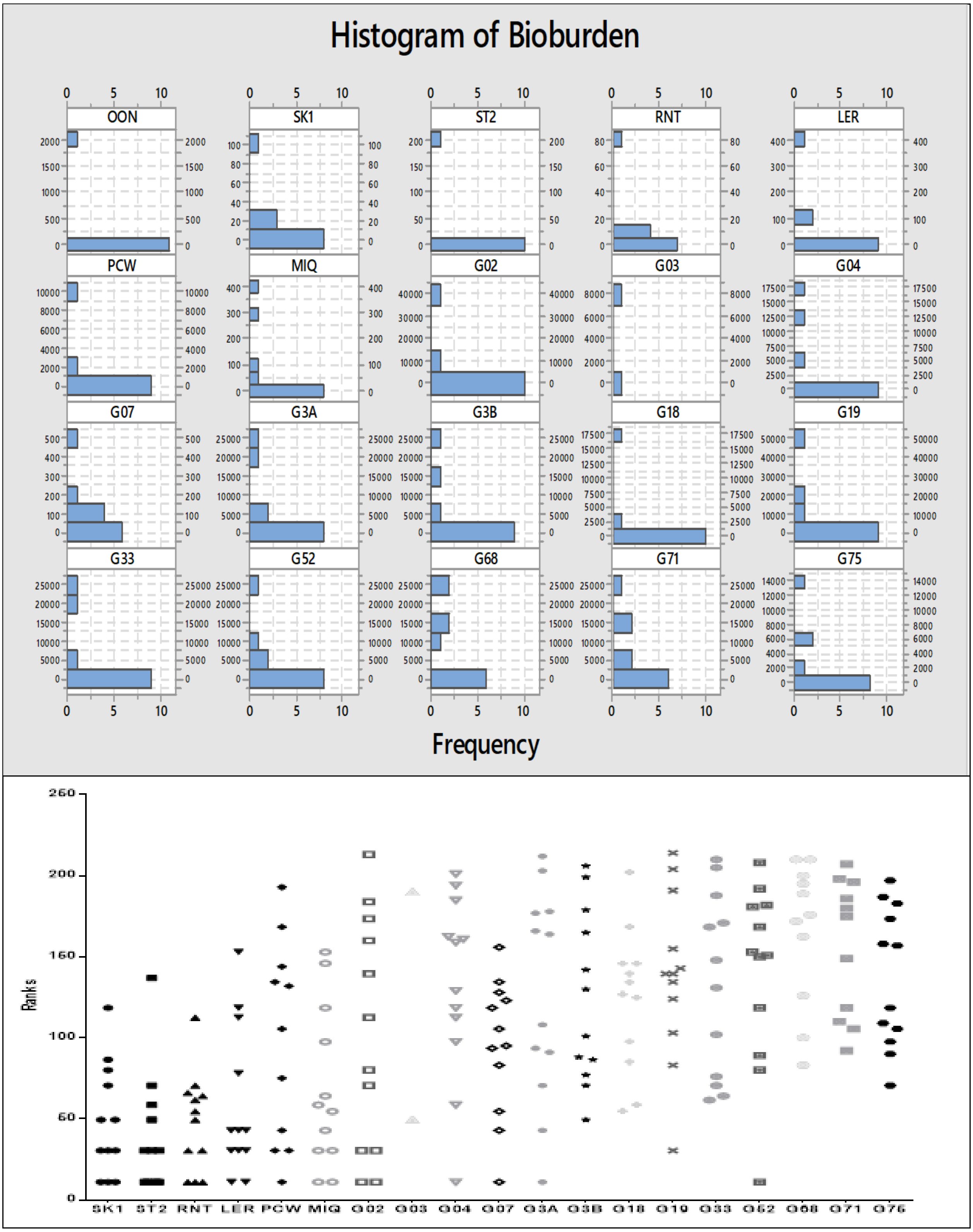

Figure 9. Histogram for distribution visualization (upper graph) and Kruskal-Wallis test showing the ranks of microbiological data for each examined use point (lower graph)

Table 2. Non-parametric Spearman correlation matrix showing two-tailed correlation coefficient (r) and *p*-values

Correlation Matrix	SK1	ST2	RNT	LER	PCW	MIQ	G02	G04	G07	G3A	G3B	G18	G19	G33	G52	G68	G71	G75
SK1		0.79	0.34	0.77	-0.33	0.59	0.31	0.24	-0.29	0.41	0.60	-0.50	0.03	0.22	0.80	0.57	0.31	0.29
ST2	0.79		0.31	0.65	-0.53	0.51	0.12	0.34	-0.33	0.23	0.47	-0.61	0.40	-0.07	0.87	0.30	0.10	0.13
RNT	0.34	0.31		0.56	0.25	-0.06	-0.22	0.23	0.17	-0.04	-0.04	-0.47	-0.19	-0.35	0.37	0.25	0.11	-0.07
LER	0.77	0.65	0.56		-0.46	0.36	0.09	0.27	0.01	0.17	0.13	-0.46	0.04	0.12	0.57	0.35	0.23	0.00
PCW	-0.33	-0.53	0.25	-0.46		-0.35	0.09	0.18	0.31	-0.14	-0.17	0.20	-0.47	-0.13	-0.41	-0.16	0.02	-0.15
MIQ	0.59	0.51	-0.06	0.36	-0.35		0.66	0.14	-0.44	0.39	0.72	0.20	0.47	0.60	0.52	0.20	-0.19	0.54
G02	0.31	0.12	-0.22	0.09	0.09	0.66		0.32	-0.24	0.18	0.48	0.33	0.32	0.67	0.17	0.03	0.12	0.15
G04	0.24	0.34	0.23	0.27	0.18	0.14	0.32		0.09	0.34	0.09	0.02	0.05	0.36	0.39	0.06	0.35	-0.16
G07	-0.29	-0.33	0.17	0.01	0.31	-0.44	-0.24	0.09		-0.43	-0.73	0.15	-0.39	-0.41	-0.57	-0.36	-0.16	-0.13
G3A	0.41	0.23	-0.04	0.17	-0.14	0.39	0.18	0.34	-0.43		0.72	0.20	-0.11	0.70	0.48	0.76	0.49	0.31
G3B	0.60	0.47	-0.04	0.13	-0.17	0.72	0.48	0.09	-0.73	0.72		-0.02	0.17	0.56	0.69	0.71	0.36	0.53
G18	-0.50	-0.61	-0.47	-0.46	0.20	0.20	0.33	0.02	0.15	0.20	-0.02		0.08	0.57	-0.57	-0.21	-0.30	0.14
G19	0.03	0.40	-0.19	0.04	-0.47	0.47	0.32	0.05	-0.39	-0.11	0.17	0.08		0.20	0.19	-0.36	-0.50	-0.29
G33	0.22	-0.07	-0.35	0.12	-0.13	0.60	0.67	0.36	-0.41	0.70	0.56	0.57	0.20		0.17	0.31	0.25	0.15
G52	0.80	0.87	0.37	0.57	-0.41	0.52	0.17	0.39	-0.57	0.48	0.69	-0.57	0.19	0.17		0.55	0.36	0.28
G68	0.57	0.30	0.25	0.35	-0.16	0.20	0.03	0.06	-0.36	0.76	0.71	-0.21	-0.36	0.31	0.55		0.74	0.45
G71	0.31	0.10	0.11	0.23	0.02	-0.19	0.12	0.35	-0.16	0.49	0.36	-0.30	-0.50	0.25	0.36	0.74		0.14
G75	0.29	0.13	-0.07	0.00	-0.15	0.54	0.15	-0.16	-0.13	0.31	0.53	0.14	-0.29	0.15	0.28	0.45	0.14	

Correlation coefficient (r)	Lower	Upper	Lower	Upper
Very weak	0.00	0.19	0.00	-0.19
Weak	0.20	0.39	-0.20	-0.39
Moderate	0.40	0.59	-0.40	-0.59
Strong	0.60	0.79	-0.60	-0.79
Very Strong	0.80	1.00	-0.80	-1.00

*p**	SK1	ST2	RNT	LER	PCW	MIQ	G02	G04	G07	G3A	G3B	G18	G19	G33	G52	G68	G71	G75
SK1		0.6	27.1	0.5	28.9	4.8	31.6	45.3	33.1	18.3	4.3	8.3	91.7	49.6	0.3	7.1	36.2	36.1
ST2	0.6		34.6	3.6	9.0	11.3	72.6	29.9	27.2	48.3	15.0	3.9	22.2	76.2	0.1	36.2	76.2	69.8
RNT	27.1	34.6		6.2	45.8	81.2	45.6	47.6	60.4	88.0	87.2	11.0	53.3	25.4	23.5	44.7	74.6	81.1
LER	0.5	3.6	6.2		13.0	25.0	77.7	39.3	97.8	60.2	69.0	12.2	91.4	71.6	5.7	29.1	48.9	96.9
PCW	28.9	9.0	45.8	13.0		27.1	78.2	60.0	35.9	66.7	60.8	55.9	13.7	69.8	20.5	62.5	95.3	63.8
MIQ	4.8	11.3	81.2	25.0	27.1		2.2	66.2	14.7	21.0	1.1	52.0	12.8	4.4	8.9	55.7	56.0	7.5
G02	31.6	72.6	45.6	77.7	78.2	2.2		30.4	44.5	57.5	11.3	28.8	30.7	2.0	59.8	92.2	71.7	64.6
G04	45.3	29.9	47.6	39.3	60.0	66.2	30.4		78.3	27.6	78.3	95.1	89.1	25.6	21.0	85.4	28.6	61.9
G07	33.1	27.2	60.4	97.8	35.9	14.7	44.5	78.3		16.9	0.9	64.6	21.1	18.4	5.6	26.8	63.4	68.3
G3A	18.3	48.3	88.0	60.2	66.7	21.0	57.5	27.6	16.9		1.1	52.3	72.8	1.4	12.1	0.9	12.9	33.1
G3B	4.3	15.0	87.2	69.0	60.8	1.1	11.3	78.3	0.9	1.1		95.1	59.9	6.3	1.6	1.7	27.3	7.9
G18	8.3	3.9	11.0	12.2	55.9	52.0	28.8	95.1	64.6	52.3	95.1		79.7	5.6	5.4	51.2	35.9	65.5
G19	91.7	22.2	53.3	91.4	13.7	12.8	30.7	89.1	21.1	72.8	59.9	79.7		53.8	55.3	26.3	12.2	35.0
G33	49.6	76.2	25.4	71.6	69.8	4.4	2.0	25.6	18.4	1.4	6.3	5.6	53.8		58.8	35.1	45.1	63.5
G52	0.3	0.1	23.5	5.7	20.5	8.9	59.8	21.0	5.6	12.1	1.6	5.4	55.3	58.8		8.3	27.3	37.9
G68	7.1	36.2	44.7	29.1	62.5	55.7	92.2	85.4	26.8	0.9	1.7	51.2	26.3	35.1	8.3		1.2	16.9
G71	36.2	76.2	74.6	48.9	95.3	56.0	71.7	28.6	63.4	12.9	27.3	35.9	12.2	45.1	27.3	1.2		69.4
G75	36.1	69.8	81.1	96.9	63.8	7.5	64.6	61.9	68.3	33.1	7.9	65.5	35.0	63.5	37.9	16.9	69.4	

* Expressed as %

p-value	p-value %	Evidence for rejecting H_0
More than 0.1	> 10%	Very weak to none
Between 0.1 - 0.05	5% 10%	Weak
Between 0.05 - 0.01	5% 1%	Strong
Less than 0.01	< 1%	Very strong

4.2. Microbiological fluctuations and outliers over study period for the examined point-of-use

The results in the corresponding result section provided highlights the significance of microbiological fluctuations and outliers at various water use ports, as well as the use of the Coefficient of Variation (CV) to assess the variability and precision of microbiological counts in a water line. The reference to Insee (2016) adds credibility to the explanation of the CV as a statistical measure that relates the standard deviation of a set of values to its mean. Furthermore, the reference to Hayes (2022) supports the assertion that the CV is used to compare the degree of variability between distinct sets of microbiological counts, reinforcing the importance of this statistical measure in the context of water quality assessment.

Additionally, the reference to Jelliffe et al. (2015) underscores the utility of the CV in assessing the precision of estimates, further emphasizing its relevance in evaluating water quality. The explanation of how a lower CV indicates a more precise estimation of water quality at certain ports, while a higher CV suggests greater uncertainty or variability, provides a clear understanding of the implications of the CV in this context. The discussion of fluctuation in the coefficient of variation as a useful statistical tool for assessing variability and precision in water distribution system datasets is supported by the references and adds depth to the analysis. Finally, the information provided offers a comprehensive understanding of the significance of microbiological fluctuations and outliers at water use ports, as well as the use of the CV as a statistical measure to assess variability and precision in water quality assessment.

4.3. Descriptive statistical distribution of the data related to microbiology

The establishment of an early warning alarm using a cut-off threshold for microbiological counts in city water distribution lines is a critical measure for safeguarding public health. The reference to Eissa et al. (2015) supports the rationale for setting a warning limit for microbiological counts below the action threshold, emphasizing the importance of this early warning system in monitoring water quality. Furthermore, the use of the non-parametric *t*-test analog test to identify lines with a higher potential of bioburden density, as mentioned in the context of GraphPad Software, LLC (2023), adds credibility to the methodology used in this study. This statistical approach is crucial for identifying segments of the distribution system that exhibit aberrant microbial density, providing a basis for establishing the early warning alarm.

The discussion of distribution segments that showed actual medians statistically not significantly different from the reference value at $\alpha=0.05$, such as use ports G68, G71, and G3A, as well as the indication of discrepancy values and two-tailed *p*-values, as shown in Figure 4, demonstrates the rigorous statistical analysis employed in identifying the sources of microbial density aberrancy. The reference to the remaining points of use being far below the alarming limit and statistically different from it further reinforces the effectiveness of the early warning alarm system in distinguishing between acceptable and aberrant microbiological counts. In summary, the information provided offers a comprehensive understanding of the rationale and statistical methodology behind the establishment of an early warning alarm using a cut-off threshold for microbiological counts in city water distribution lines.

4.4. Comparing microbiological record data from under investigation points of use pairwise

The significance of variations between different municipal water distribution segments can be investigated through non-parametric multiple comparisons using the Kruskal-Wallis test, as indicated by Eissa (2016b). This statistical approach is crucial in the medical and public health context for understanding and addressing variations in bioburden quality across different lines of the distribution system. The reference to Table 1, which highlights the main findings of the statistical test, provides empirical evidence of detectable differences in bioburden quality between different lines of the distribution system, underscoring the importance of investigating these variations.

Furthermore, the detailed examination of the sources of significant variability projected from Figures 6 and 7, which depict the mean rank difference for significant and non-significant pairwise comparisons, respectively, offers a comprehensive understanding of the nature and extent of the variations in bioburden quality. The identification of a few use points as the reason behind the significance of these variations, due to their contribution to unusually high microbial density in water samples, is a critical insight that can inform targeted interventions to address the sources of contamination. The main effectors contributing to unusually high microbial density in water samples are the factors marked with significance levels (*, **, ***), which indicate that there are statistically significant differences in the mean rank of the bacterial counts among the different use points. These factors could include contamination sources, environmental conditions, or human activities affecting water quality. For example, some of the possible factors are: type of water source, implemented treatment, location of the site, population growth, lack of protection, agriculture, urbanization/sanitation, and flooding threats.

These factors affect the overall water quality by introducing or promoting the growth of pathogenic or opportunistic microorganisms, such as Escherichia coli, Legionella, Pseudomonas, etc., which can pose health risks to humans and animals. Moreover, these factors can also alter the physicochemical properties of water, such as pH, turbidity, dissolved oxygen, etc., which can affect the taste, odor, color, and clarity of water.

The reference to Figure 8, which illustrates the count of the main effectors in the municipal distribution system in relation to other stable and low bioburden water lines, further supports the identification of key points of discrepancy and their impact on overall bioburden quality. Additionally, the reference to Figure 5, which identifies points of discrepancy between two distinct groups, and Figure 9, which provides criteria for ranking the microbiological sampling output results and demonstrates the range, spreading, and scatter of the data points, offers valuable insights into the distribution and shape of microbial count data. In summary, the information provided offers a comprehensive understanding of the significance of variations in bioburden quality in municipal water distribution segments and the importance of employing non-parametric statistical tests and visualizations to investigate and address these variations from a medical and public health perspective.

4.5. Non-parametric correlation matrix for the municipal water system's points of use

The significance of variations between different municipal water distribution segments could be investigated through non-parametric multiple comparisons Kruskal-Wallis test (Eissa, 2016b). The apparent shape of the distribution in the histograms of Figure 8, showing truncated spreading of the microbial count data, is a crucial observation from a public health and medical perspective. This observation suggests that the datasets do not follow a normal distribution, highlighting the non-normal nature of the microbial count data. The surveying of the correlation between use points under examination using a non-parametric Spearman matrix, as conducted by Eissa and Rashed (2022), is a significant approach in understanding the relationships and associations within the microbial count data. The detailed output presented in Table 2, which includes the calculated two-tailed correlation coefficient (r) and the corresponding *p*-values, provides valuable insights into the strength of the associations and correlations among the use points.

The identification of the lowest negative but strong correlation between G07 and G3B, along with the visual color marking for ease of scanning and spotting of the significant correlation, underscores the importance of understanding the interrelationships between different use points in the distribution system. The specific correlations identified, such as the strong associations between G02 and PCW, SK1 and ST2, G52 and G68, and others, offer valuable insights into the interconnectedness of the use points and their impact on the bioburden data. The presence of very strong positive correlations between G52 and ST2, as well as strong correlations between G71 and G68, further highlights the complex interdependencies within the bioburden data, providing critical information for public health and medical interventions. In summary, the findings from the non-parametric Spearman matrix analysis provide valuable insights into the correlations and associations between use points in the distribution system, offering a comprehensive understanding of the relationships within the microbial count data from a public health and medical perspective.

The significance of variations between different municipal water distribution segments can be investigated through non-parametric multiple tests. This statistical approach is crucial in the medical and public health context for understanding and addressing variations in microbial quality across different lines of the distribution system. The presence of very strong positive correlations between few points within the municipal network system, highlight the complex interdependencies within the bioburden data and provide critical information for public health and medical interventions. The identification of a few use points as the reason behind the significance of these variations, due to their contribution to unusually high microbial density in water samples, is a critical insight that can inform targeted interventions to address the sources of contamination. Finally, the non-parametric statistical tests reveal significant variations in bioburden quality between different use points and lines of the distribution system, which have important implications for water quality monitoring and public health protection strategy.

5. Conclusion

The biological stability of the municipal distribution systems is an important aspect to attain and maintain. Monitoring and controlling the microbiological quality throughout the distribution lines is vital until reaching the final consumers. Only six points from the 19 water-use ports showed more than 65% of the total bioburden recovered from this study with one line alone demonstrating an average of 16%. The study is limited by the amount of data available at the early stages of the infrastructure examination. More results are required to be gathered to construct control charts that might show the behavior of the inspection characteristics under investigation. However, the municipal system showed at the initial examination that control measures must be put in place to improve the microbiological stability despite the fact of presence of low microbial count within some sections in the system yet the fluctuations and variabilities were considerably high. Correlating obtained results with layout infrastructure should be addressed in other separate investigations using engineering design to correct for structure-related issues. Future research could build upon the findings of this study by expanding the scope and frequency of data collection, as well as exploring the factors that influence the biological stability of the municipal distribution systems.

Future research can refine our understanding of bioburden variations in water distribution systems. Collecting more samples from diverse water use points over extended periods would allow for more accurate control charts and trend identification. Examining the connections between biological instability and water source, treatment, pipe materials, and flow rates could shed light on root causes. Finally, experimenting with interventions like treatment modifications, pipe cleaning, or biostatic agents would help determine the most effective strategies for reducing bioburden and ensuring consistent water quality.

Maintaining the biological stability of municipal water distribution systems is crucial for public health. Our study highlights that only six out of 19 water-use ports contribute significantly to the total bioburden, warranting immediate control measures. While data limitations exist, this work lays the groundwork for future investigations, emphasizing the need for control charts and correlation studies with infrastructure design. The inconclusive correlation matrix suggests potential design-related defects, prompting further exploration using logarithmic transformations. Recommendations include investigating the impact of source feed water and advocating for regular monitoring programs nationwide. The presented statistical analysis by public health authorities remains pivotal for ongoing assessments and enhancements in water safety protocols.

Conflict of interest

The author declares that there is no conflict of interest.

Ethical Approval

For this type of study, formal consent is not required.

References

Apure (2022). What is municipal water? https://apureinstrument.com/blogs/what-is-municipal-water/

Bruyninckx, H. (2018). Clean water is life, health, food, leisure, energy. https://www.eea.europa.eu/signals/signals-2018-content-list/articles/clean-water-is-life-health

Cabral, J. P. S. (2010). Water Microbiology. Bacterial Pathogens and Water. *International Journal of Environmental Research and Public Health*, *7*(10), 3657-3703. https://doi.org/10.3390/ijerph7103657

CDC (2020). Importance of water quality and testing. https://www.cdc.gov/healthywater/drinking/public/water_quality.html

CDC (2022). Water treatment. https://www.cdc.gov/healthywater/drinking/public/water_treatment.html

Dinka, M. O. (2018). Safe drinking water: Concepts, benefits, principles and standards. In M. Glavan (Ed.), *Water Challenges of an Urbanizing World*. IntechOpen. https://doi.org/10.5772/intechopen.71352

Edam, W., & Abdelgalil, E. E. (2022). Seasonal Quality of Water Sources Indicated by Escherichia coli: A Case of Elobied, North Kordofan State, Sudan. *Applied Water Science*, *12*, 50. https://doi.org/10.1007/s13201-022-01578-7

Eissa, M. E. (2015). Shewhart Control Chart in Microbiological Quality Control of Purified Water and Its Use in Quantitative Risk Evaluation. *Pharmaceutical and Bioscience Journal*, *4*(1), 45-51.

Eissa, M. (2016a). Study of Microbial Distribution from Different Processing Stages in Purified Water Production Plant of Pharmaceutical Manufacturing Facility. *Research and Reviews: Journal of Microbiology and Virology*, *6*(1), 31-45.

Eissa, M. E. (2016b). Extended Application of Dose-response Models of Infection on Injectable Insulin Products in Vials. *Russian Open Medical Journal*, *5*(3), e0305. https://doi.org/10.15275/rusomj.2016.0305

Eissa, M. E. (2017). Determination of the Microbiological Quality of Feed City Water to Pharmaceutical Facility: Distribution Study and Statistical Analysis. *Athens Journal of Sciences*, *4*(2), 143-160. https://doi.org/10.30958/ajs.4-2-4

Eissa, M. E. (2018). Investigation of Microbiological Quality of Water from the Feed Source to the Terminal Application in the Healthcare Facility: A Case Study. *Health Research*, *2*(1), 16-23. https://doi.org/10.31058/j.hr.2018.21002

Eissa, M., & Rashed, E. (2022). Descriptive Analysis of Coronavirus Disease Cases based on Geographical Distribution in Canadian Provinces/Territories: Statistical Investigation into Epidemiological Pattern. *Academia Letters*, Article 5191. https://doi.org/10.20935/AL5191

Eissa, M., Eissa, D., & Rashed, E. (2022). Principal Component Analysis in Long Term Assessment of Total Viable Plate Count of Municipal Water Distribution Network System in Healthcare Facility. *Environmental Research and Technology*, *5*(2), 165-171. https://doi.org/10.35208/ert.1062683

Eissa, D. E., Rashed, E. R., & Eissa, M. E. (2022). Dendrogram Analysis and Statistical Examination for Total Microbiological Mesophilic Aerobic Count of Municipal Water Distribution Network System. *HighTech and Innovation Journal*, *3*(1), 28-36. https://doi.org/10.28991/HIJ-2022-03-01-03

Eissa, M. E., Seif, M., & Fares, M. (2015). Assessment of Purified Water Quality in Pharmaceutical Facility using Six Sigma Tools. *International Journal of Pharmaceutical Quality Assurance*, *6*(2), 54-72.

Figueras, M. J., & Borrego, J. J. (2010). New Perspectives in Monitoring Drinking Water Microbial Quality.

International Journal of Environmental Research and Public Health, *7*(12), 4179-4202. https://doi.org/10.3390/ijerph7124179

GraphPad (2023). GraphPad Prism 9 user guide. https://www.graphpad.com/guides/prism/latest/user-guide/index.htm

Hayes, A. (2022). Co-efficient of variation meaning and how to use it. https://www.investopedia.com/terms/c/coefficientofvariation.asp

Hayward, C., Ross, K. E., Brown, M. H., Bentham, R., & Whiley, H. (2022). The Presence of Opportunistic Premise Plumbing Pathogens in Residential Buildings: A Literature Review. *Water*, *14*(7), 1129. https://doi.org/10.3390/w14071129

Health Canada (2021). Overview of the microbiological aspects of drinking water quality. https://www.canada.ca/en/health-canada/services/environmental-workplace-health/reports-publications/water-quality/guidance-document-overview-microbiological-aspects-drinking-water-quality.html

Hibbing, M. E., Fuqua, C., Parsek, M. R., & Peterson, S. B. (2010). Bacterial Competition: Surviving and Thriving in the Microbial Jungle. *Nature Reviews Microbiology*, *8*(1), 15-25. https://doi.org/10.1038/nrmicro2259

Ibekwe, A. M., & Murinda, S. E. (2019). Linking Microbial Community Composition in Treated Wastewater with Water Quality in Distribution Systems and Subsequent Health Effects. *Microorganisms*, *7*(12), 660. https://doi.org/10.3390/microorganisms7120660

Insee (2016). Definition - Coefficient of variation / CV. https://www.insee.fr/en/metadonnees/definition/c1366

Jelliffe, R. W., Schumitzky, A., Bayard, D., Fu, X., & Neely, M. (2015). Describing Assay Precision - Reciprocal of Variance is Correct, not CV Percent: Its Use Should Significantly Improve Laboratory Performance. *Therapeutic Drug Monitoring*, *37*(3), 389-394. https://doi.org/10.1097/FTD.0000000000000169

Liu, G., Zhang, Y., van der Mark, E., Magic-Knezev, A., Pinto, A., van den Bogert, B., Liu, W., van der Meer, W., & Medema, G. (2018). Assessing the Origin of Bacteria in Tap Water and Distribution System in an Unchlorinated Drinking Water System by SourceTracker using Microbial Community Fingerprints. *Water Research*, *138*, 86-96. https://doi.org/10.1016/j.watres.2018.03.046

Netsol Water Solutions Pvt (2022). What is the difference between well water and municipal water. https://www.netsolwater.com/how-is-well-water-different-from-city-municipal-water.php?blog=934

Noiva, K., Fernández, J. E., & Wescoat, J. L. (2016). Cluster Analysis of Urban Water Supply and Demand: Toward Large-scale Comparative Sustainability Planning. *Sustainable Cities and Society*, *27*, 484-496. https://doi.org/10.1016/j.scs.2016.06.016

Park, J. W., Kim, H. C., Meyer, A. S., Kim, S., & Maeng, S. K. (2016). Influences of NOM Composition and Bacteriological Characteristics on Biological Stability in a Full-scale Drinking Water Treatment Plant. *Chemosphere*, *160*, 189-198. https://doi.org/10.1016/j.chemosphere.2016.06.078

Prest, E. I., Hammes, F., van Loosdrecht, M. C. M., & Vrouwenvelder, J. S. (2016). Biological Stability of Drinking Water: Controlling Factors, Methods, and Challenges. *Frontiers in Microbiology*, *7*, 45. https://doi.org/10.3389/fmicb.2016.00045

Pruden, A. (2014). Balancing Water Sustainability and Public Health Goals in the Face of Growing Concerns about Antibiotic Resistance. *Environmental Science and Technology*, *48*(1), 5-14. https://doi.org/10.1021/es403883p

Sensorex (2020). Municipal water: How it works, treatment, and applications. https://sensorex.com/2020/03/10/municipal-water-treatment-applications/

US EPA (2013). Regulatory and guidance information by topic: Water. https://www.epa.gov/regulatory-information-topic/regulatory-and-guidance-information-topic-water

Vacs Renwick, D., Heinrich, A., Weisman, R., Arvanaghi, H., & Rotert, K. (2019). Potential Public Health Impacts of Deteriorating Distribution System Infrastructure. *Journal American Water Works Association*, *111*(2), 42-53. https://doi.org/10.1002/awwa.1235

World Health Organization (2008). Guidelines for drinking-water quality [electronic resource]: Incorporating 1st and 2nd addenda, Vol. 1, recommendations (3rd ed.). https://apps.who.int/iris/handle/10665/204411

World Health Organization (2011). Guidelines for drinking-water quality (4th ed.). https://apps.who.int/iris/handle/10665/44584

RESEARCH ARTICLE
Journal of Biometry Studies (2024) 4(1): 31-41
DOI: 10.61326/jofbs.v4i1.03

www.biometrystudies.com

Journal of Biometry Studies

PRENSIP
www.prensip.gen.tr

Evaluation of the effect of morphological traits on fish growth by comparison using ridge and ordinary least squares regression

Mehmet Fatih CAN[1,*], Cemil KARA[2]

[1]İskenderun Technical University, Faculty of Marine Science and Technology, Hatay/TÜRKİYE
[2]Karadeniz Technical University, Faculty of Science, Department of Biology, Trabzon/TÜRKİYE

[*]Corresponding author: mfatih.can@iste.edu.tr

Received: 26/04/2024, Accepted: 21/06/2024

Abstract

Growth in fish is characterized by length and weight and studies that encompassing relationships among the fish length and weigh with some morphometric traits provide crucial information in the field of fish biology. High correlations (or multicollinearity) among the morphometric traits in fish morphology studies is a well-known phenomenon. If the relationship between growth and morphometry is modeled using the ordinary least squares estimator (OLS), the parameter estimates are likely to be too large in absolute value and possibly have the wrong sign due to the problem of multicollinearity. The ridge regression (RR) estimator has been proposed to avoid the adverse effects of multicollinearity among the regressors. In this study, therefore, multiple linear RR was used to model relationship fish length and fish weight to some predictive metric traits. Predictive traits were predorsal length, head length, post dorsal length, head height, and eye diameter. Data were derived from a total of 126 *Capoeta damascina* individuals sampled from the Euphrates, Türkiye. The ridge optimal k-parameter, producing acceptable variance influence factor (VIF<10), was determined as 0.021 for both of growth indices considering a combination of ridge trace and VIF trace plots and MSE and VIF values produced from an array of k values (0.0≤ k≤1.00). Although, small decreases were observed in the adjusted R-square values (ARS) obtained by RR compared to the ARS values obtained by OLS, a significant decrease in VIF values outweighs this drawback, indicating that the models are more stable. The methodological approach and findings in this study may contribute to filling a gap in the literature regarding the relationships between fish growth and morphology. Additionally, it could enhance better growth predictions for different fish species, aiding sustainable fisheries management and the selective cultivation of desired fish traits in aquaculture by improving the understanding of morphometric features.

Keywords: Fish morphometry, Multicollinearity, Variance influence factor, *Capoeta damascina*

Please cite this article as follows:

Can, M. F., & Kara, C. (2024). Evaluation of the Effect of Morphological Traits on Fish Growth by Comparison Using Ridge and Ordinary Least Squares Regression. *Journal of Biometry Studies*, *4*(1), 31-41. https://doi.org/10.61326/jofbs.v4i1.03

1. Introduction

Fish morphometric traits are quantifiable characteristics related to the shape, size, and proportions of fish bodies. Morphometric studies encompassing these relationships provide crucial information in the field of fish biology, including species growth, feeding behavior, ecological strategies, niche partitioning, habitat utilization, and trophic structure. For instance, examining the linear relationships between total length, body height, and width reveals growth patterns, the height of the caudal peduncle influences swimming efficiency and overall fitness, while pectoral fin length and eye diameter reflect adaptations related to habitat, feeding, and locomotion. These morphometric analyses contribute significantly to our understanding of fish diversity and ecological dynamics (Clabaut et al., 2007; Mojekwu & Anumudu, 2015; Keppeler et al., 2020; Pino et al., 2021). In addition to these, morphometric features can also be used in aquaculture to distinguish between fish escaped from the rearing environment and wild fish (Arechavala-Lopez et al., 2012; Lenhardt et al., 2012).

Morphometric characteristics are taken into account in stock discrimination studies in fisheries science using multivariate statistical methods such as Principal Component Analysis, Cluster Analysis, and Linear Discriminant Analysis (Winans, 1987; Turan, 1999; Mojekwu & Anumudu, 2015). But, due to allometric growth that is an inherent characteristic of fishes which is a common problem with morphometric data is that all measurements are highly correlated with length, the morphometric traits are required to be transformed before analysis in stock discriminant studies. This high correlation between morphometric features is called multicollinearity. Due to this situation, the simple linear regression method is generally used to evaluate the relationships between fish growth and morphometric characteristics, except for stock discrimination studies (Kyritsi & Moutopoulos, 2018; Orbach et al., 2019; Yazıcı & Yazıcıoğlu, 2020).

In cases where there is multicollinearity; (i) The values of parameters estimated by the Ordinary Least Squares (OLS) method may differ significantly from their actual values, (ii) Regression coefficients are uncertain and the standard errors of these coefficients are infinite, (iii) Variance and covariances of the regression coefficients are tend to increase, (iv) *R*-square value of the model is high, However, none or very few of the independent variables may be significant according to the partial *t*-test, and (v) The direction of the relationships of the relevant independent variables with the dependent variable may contradict theoretical and experimental expectations. Although there are statistical approaches such as principal component analysis (PCA) and removing the variable having VIF higher than 10 are used to overcome this problem, the use of these approaches may lead to information loss. Another approach to overcome this problem is to use ridge regression. Ridge regression is a statistical method used to estimate coefficients in multiple regression models when independent variables are highly correlated without removing any variables from model (Albayrak, 2005; Karakaş, 2008; McDonald, 2009; Büyükuysal & Öz, 2016; Nandi & Saikia, 2015; Özkale & Altuner, 2021).

Although ridge regression has been used for different purposes in fish related studies (Rikardsen & Johansen, 2003; Okamura et al., 2017; Niu et al.,2019; Martins et al., 2021; Liyandja et al., 2022; Gao et al., 2024), it seems that it has not been used to describe the relationship between fish morphology and fish growth to date. In essence, the fish material used in this study were considered for example purposes only to illustrate why and how ridge regression is used in modeling the relationship between fish growth and morphometry. Therefore, the main purpose of this study is to fill a gap in the literature on this subject by holistically estimating the relationship between total fish length and total weight characteristics and a group of morphometric features using the ridge regression method. Moreover, the methodological approach in the study can contribute to (i) sustainable fisheries management by making better growth predictions for different fish species and (ii) a better understanding.

2. Material and Methods

2.1. Fish material

The morphological features used in this study was belong to the *Capoeta damascina* species. The specimen measured in this study were provided from a research project that was supported by the Adıyaman University Scientific Research Projects (BAP) Coordination Unit, Project No. 2012/001, and was carried out with the legal consent of the Ministry of Food, Agriculture, and Livestock, General Directorate of Fisheries and Aquaculture (permission date 05.04.2012; permission number: 01515)**.** The *C. damascina* species group occurs in the entire Levant, Mesopotamia, the Orontes, Iran, and the southern and eastern parts of Türkiye (Alp et al., 2005; Alp et al., 2013; Asadollah et al., 2017). A total of 126 *C. damascina* specimens caught from the Adıyaman region (in the middle Euphrates Basin) of Türkiye between April 2012 and December 2013 were caught in streams by using electroshock devices. Then, all length related measurements (mm) were made with a digital caliper with a precision of 0.01 mm, and the weight measurements were made with a digital scale with a precision of 0.01 g. Since this study was based on a modeling methodology approach, there was no need to provide further information about the fish and the area where they were sampled.

2.2. Metric traits

In the study total lenght (TL), total weight (TW), body height (BH), predorsal length (PR), head length (HL), post dorsal length (PD), head height (HH) and eye diameter (ED) metric features were taken into consideration.

2.3. Statistical framework

2.3.1. The general framework of the study

In the study total lenght (TL), total weight (TW), body height (BH), predorsal length (PR), head length (HL), post dorsal length (PD), head height (HH) and eye diameter (ED) metric features were taken into consideration.

Irrespective of used domain, the biggest debate with the ridge estimator is the selection of the regularization parameter (*k*). Therefore, the proposed general framework of the study was given below in hierarchical format:

a. Calculation of some descriptive statistics of all variables

b. Calculation and visualization of the correlation matrix between variables: It will give us a preliminary idea about whether there is a multicollinearity problem.

c. Calculation of VIF values for predictor variables based on Ordinary Least Square (OLS): If there is a doubt determined in step “b”, VIF values, which are indicative of multicollinearity, are calculated for each predictor variable. The presence of variables with VIF equal and greater than 10 quantitatively indicates the presence of multicollinearity.

d. To have an idea about the approximate optimum k-parameter visually using graphical methods (VIF Trace and Ridge Trace Plot) for a range of the Ridge k parameter.

e. Calculate the models' minimum Mean Square Error (MSE), minimum generalized cross-validation (GCV) and adjusted *R*-squared values for a range of Ridge k parameters.

f. Calculate the VIF values of the predictor variables for a range of the ridge k-parameter.

g. Deciding on the optimum ridge k value by considering the smallest MSE, adj-*R*-square and VIF values in a balanced manner.

2.3.2. Models

In this study TL and TW were considered dependent and other metric traits were considered independent (regressor) variables. Regardless relationship between dependent and independent variables, to simplify the modelling dependent variables were modelled with multiple linear regression approach as following:

$$TL = \alpha + \beta_1 PR + \beta_2 PD + \beta_3 HH + \beta_4 BH + \beta_5 HL + \beta_6 ED + \varepsilon$$

$$TW = \alpha + \beta_1 PR + \beta_2 PD + \beta_3 HH + \beta_4 BH + \beta_5 HL + \beta_6 ED + \varepsilon$$

where, α, β, and ε are intercept, regression coefficients, and error term, respectively.

2.3.3. Multicollinearity and variance inflation factor (VIF)

In regression analysis, some assumptions are taken into account when estimating the parameters that show the relationship between dependent and independent variables. In order to determine the relationship between variables with multiple regression analysis, the assumptions of the multiple linear regression model must be met. However, the assumptions of the multiple regression model are not valid in the analysis of every event. The assumptions in question are the assumptions about the error vector and the independent variable matrix. One of these assumptions that ensures that the standard errors of the estimates are smallest and therefore the parameter estimates are effective is that there is no relationship between the independent variables. This is called multicollinearity. As can be seen from the graphs above (Figure 1 and Figure 2), the existence of high correlations between the predictor variables is a sign of this situation. If there is a perfect correlation between the explanatory variables, that is, if the correlation coefficient for these variables is equal to 1, the parameters become undeterminable. It becomes impossible to find separate numerical values for each parameter, and in this case the least squares method cannot be used. On the other hand, if there is no correlation between the explanatory variables, that is, if the correlation coefficient for these variables is equal to 0, they are called orthogonal variables and do not pose a problem in estimating the coefficients. Multicollinearity may occur for various reasons (Katz, 2006; Adeboye et al., 2014; Daoud, 2017). However, considering the current study, the main reason for multicollinearity is that the variables of interest tend to change together over time.

One of the approaches used to detect multicollinearity is the variance inflation factor (VIF). The variance inflation factor is calculated to determine the degree of relationship of an independent variable with other independent variables. If there is no relationship between the dependent variable and the independent variables (*R*-square=0), the variance inflation factor is equal to 1 (1/(1- *R*-square)=1/(1-0)=1). If there is a perfect relationship between the dependent and independent variables (*R*-square=1), the variance inflation factor [VIF=1/(1- *R*-square)=1/(1-1)] will be infinite. If *R*-square=90%, the variance inflation factor is obtained as 10 [VIF=1/(1-0.90)=10]. In general, the following general rule applies for interpretation: if the variance inflation factor is equal to or greater than 10 (VIF>10), there is a multicollinearity problem (Thompson et al., 2017; Oke et al., 2019). In case of collinearity, it is necessary to eliminate it in order to reduce the standard error of coefficient estimates that determine the relationship between dependent and independent variables and to make more consistent estimates. One of the methods used in case of multicollinearity is Ridge regression.

2.3.4. Ridge regression

The solution technique of ridge regression (RR) is similar to the ordinary least squares (OLS) method. The RR method, is performed by adding a small and positive constant to the diagonal elements of the $(X^T X)$ matrix formed by the variables in standard form before calculating the regression coefficient estimates. Accordingly, the ridge regression solution is as follows;

$$\hat{\beta}_{RR} = (X^T X + kI)^{-1} X^T Y$$

In the formula "I" and "k" represent the identity matrix and the ridge parameter, respectively. The difference of the RR method from OLS is the presence of the k-ridge parameter. Since the ridge solution for k=0 is equivalent to the OLS solution, the ridge estimate can also be expressed as a linear transformation of the least squares estimate. In the RR method, the steps followed in the least squares method are repeated more than once. Among the parameter estimates calculated for each k that has a value between 0 and 1, those that meet the sought criterion are determined.

One of the main obstacles in using RR is in choosing an appropriate value of k. The inventors of RR suggested using a graphic which they called the ridge trace. This plot shows the RR coefficients as a function of *k*. When viewing the ridge trace, the analyst picks a value for k for which the regression coefficients have stabilized. Often, the regression coefficients will vary widely for small values of k and then stabilize. Choose the smallest value of k possible (which introduces the smallest bias) after which the regression coefficients seem to remain constant. Note that increasing k will eventually drive the regression coefficients to zero.

2.4. Analysis tools

In the study, the packages “lmridge” (Ullah et al., 2018), “psych” (Revelle & Revelle, 2015), and “rattle” (Williams, 2011) in the R program were used.

lmridge contains functions related to fitting of the RR model and provides a simple way of obtaining the estimates of RR coefficients, testing of the ridge coefficients, and computation of different ridge related statistics, which prove helpful for selection of optimal biasing parameter *k*. The four arguments of lmridge() function are described in Table 2 (Ullah et al., 2018). The syntax of default function is,

lmridge (formula, data, scaling = ("sc","scaled","centered"),k,...)

Table 1. Description of lmridge() function arguments

Argument	Description
Formula	Symbolic representation for RR model of the form, response ~ predictors.
Data	Contains the variables that have to be used in RR model.
k	The biasing parameter, may be a scalar or vector. If a *k* value is not provided, *k*=0 will be used as the default value, i.e., the OLS results will be produced.
Scalling	The methods for scaling the predictors. The sc option uses the default scaling of the predictors in correlation form; the *scaled* option standardizes the predictors having zero mean and unit variance; and the *centered* option centers the predictors.

There are many scientific approaches to choosing the correct ridge parameter. The ultimate choice of *k* for a particular application involving linear explanatory variables still remains part art, part science (Forrester & Kalivas, 2004; Khalaf & Shukur, 2005; McDonald, 2009; Mami et al., 2021). In this study, in deciding on the *k* parameter, a series of *k* parameter ($0.0 \leq k \leq 1.0$) was considered into iterative manner in each of steps described under general framework of the study.

3. Results

3.1. Descriptive statistics and correlations

Some statistics (mean, standard deviation, coefficient of variation) of the metric features taken into account in the research are given in Table 1. Cross-correlations between variables, histogram distributions of the variables, and Correlation dendrogram are given in Figure 1 and Figure 2, respectively. In terms of variability (measured by Cv, %), TW had the highest (84.09%) as a nature of weight, others somehow were similar except ED (14.97%).

Table 2. Mean, standard deviation (Sd) and coefficient of variation (CV, %) values of the variables

Metric trait	Mean	Sd	CV, %
ED	6.13	0.92	14.97
BH	34.96	8.82	25.22
HH	21.70	6.10	28.12
HL	32.48	8.07	24.86
PD	54.38	16.15	29.71
PR	68.61	17.04	24.84
TL	171.16	45.04	26.32
TW	58.58	49.26	84.09

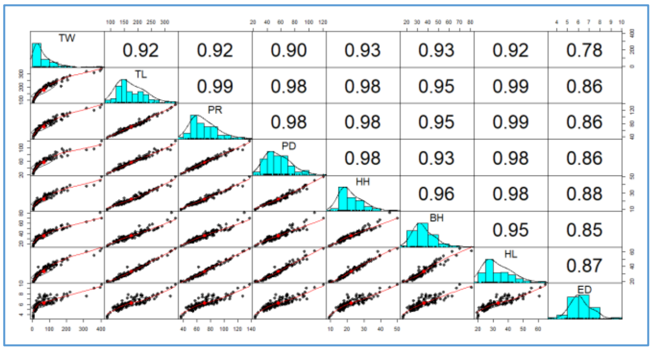

Figure 1. Cross correlations and linear relationships between variables, and histogram distribution of each variable

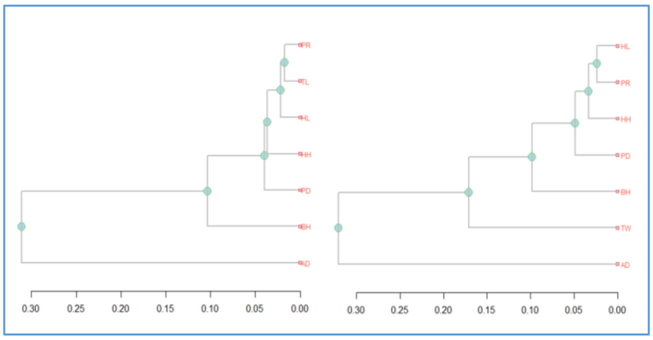

Figure 2. Correlation dendrogram graphs of variables (Total height (TL) on the left-hand side and total weight (TW) on the right-hand side)

As can be seen in Figure 1 and Figure 2, there are quite high correlations between the predictor variables (ED, BH, HH, HL, PD, and PR), and between predictors and dependent variables (TL and TW). High correlations observed between independent variables indicate a potential multicollinearity problem. However, since simple correlation coefficients alone are not sufficient, it is not possible to be sure of the existence of a multicollinearity problem (Ölmez & Yağanoğlu, 2023).

3.2. Ordinary Least Square (OLS) results and VIF values

Ridge solution for k=0 is equivalent to the ordinary least squares solution (OLS). In case k=0 for TL, the RR analysis resulted that adj-R-squared=0.984, MSE=8626.853, and regression coefficients of PR, PD, BH, and HL were significant (p<0.05), whereas HH and ED were not (p>0.05). For TW, adj-R-squared=0.886, MSE=127826.47, and only regression coefficients of BH and ED were significant (p<0.05), in contrast other predictors were not (p>0.05) (Figure 3, Figure 4).

```
Coefficients: for Ridge parameter K= 0
           Estimate Estimate (Sc) StdErr (Sc) t-value (Sc) Pr(>|t|)
Intercept   -4.0265   -31703.5376   4563.6421      -6.9470   <2e-16 ***
PR           0.9487      212.8959     49.6621       4.2869   <2e-16 ***
PD           0.7481      155.0240     38.3154       4.0460   0.0001 ***
HH          -0.9275      -75.0313     43.1890      -1.7373   0.0848 .
BH           0.4487       54.4057     24.6356       2.2084   0.0291 *
HL           2.2365      237.3432     45.1410       5.2578   <2e-16 ***
ED           0.1902        2.2332     13.5112       0.1653   0.8690
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

     R2    adj-R2  DF ridge          F       AIC        BIC
0.98540   0.98480   5.99995 1391.39510 487.13857 1137.12311

             PR       PD       HH       BH       HL      ED
 k=0   60.84427 36.21742 46.01677 14.97258 50.27037 4.50356
```

Figure 3. Regression summary and VIF values of RR for k=0 that equals to results of ordinary least squares for TL relation to predictors

```
Coefficients: for Ridge parameter K= 0
           Estimate Estimate (Sc) StdErr (Sc) t-value (Sc) Pr(>|t|)
Intercept -144.7147   -29911.3869  17566.9195      -1.7027   0.0911 .
PR           0.9802      219.9666    191.1652       1.1507   0.2521
PD          -1.0923     -226.3406    147.4884      -1.5346   0.1274
HH           3.0434      246.1986    166.2483       1.4809   0.1412
BH           3.0943      375.2203     94.8304       3.9568   0.0001 ***
HL           2.4827      263.4704    173.7621       1.5163   0.1320
ED          -9.2464     -108.5916     52.0088      -2.0879   0.0389 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Ridge Summary
        R2    adj-R2  DF ridge         F       AIC        BIC
   0.89080   0.88640   5.99995 168.53505 837.59178 1487.57632

                PR       PD       HH       BH       HL      ED
   k=0    60.84427 36.21742 46.01677 14.97258 50.27037 4.50356
```

Figure 4. Regression summary and VIF values of RR for k=0 that equals to results of ordinary least squares for TW relation to predictors

VIF values of predictor variables of both TL and TW based on ordinary least squares (OLS) were calculated as; PR=60.84, PD=36.22, HH=46.02, BH=14.97, HL=50.27, and ED=4.50. As suspected above from correlation results, VIF values clearly show that there was multicollinearity among the predictor variables for both of TL and TW (Figure 3 and Figure 4).

3.3. Searching for optimal "k" value

3.3.1. Graphical Searching

In graphical searching, the ridge trace plots and VIF trace plot for a range of k parameter (0.0≤k≤1.0) for TL and TW estimations were given in Figure 5, respectively. Ridge Trace Plot shows the effect of different k-ridge values on the regression coefficient of each predictor variable for dependent variable (TL and TW). VIF Trace plot shows the effect of different k-ridge values on the VIF values of each predictor variable for the dependent variable (TL and TW). As can be seen from the "Ridge Trace Plot" graphs for TL and TW, as the k values increase, the "Ridge" beta coefficients get closer to each other. On the graph, it is reported that the smallest MSE (7131.745) value for TL was reached at k=0.003. For TW, the lowest MSE value (59893.122) was reached at k=0.016.

VIF Trace plot shows the decrease in VIF values of predictor variables despite increasing k values according to generalized cross-validation or GCV criterion (Golub et al., 1979). The minimum GCV value for TL was reached at k=0.003, and the minimum GCV value for TW was reached at k=0.018.

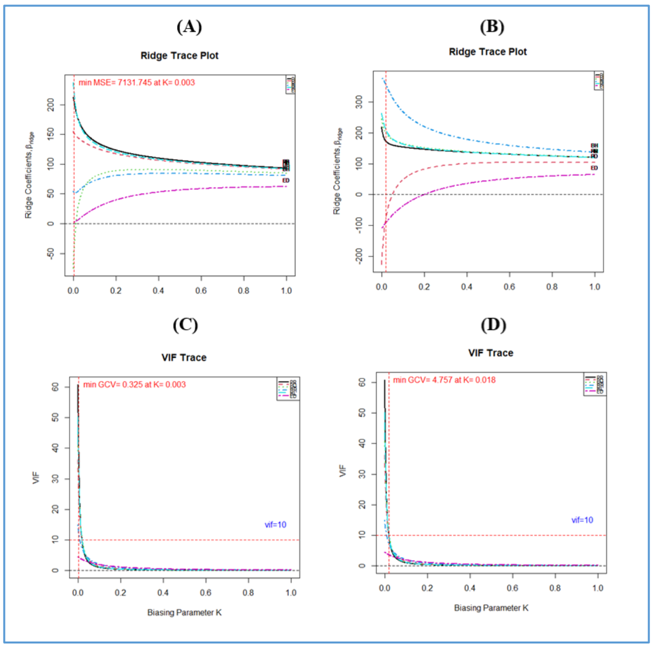

Figure 5. Ridge trace plots {TL-(A), TW-(B)} and VIF trace plot {TL-(C), TW-(D)} for a range of k parameter (0.0≤k≤1.0) for TL and TW estimations

3.3.2. Searching of the best k by minimum MSE

According to the minimum MSE approach, the ridge k value that produces the least MSE value (7131.745) for TL prediction is 0.003 and the adj-R-square value of the model is 0.9825 (Figure 6), while for TW prediction these values are found to be 59893.12, 0.016 and 0.8722, respectively (Figure 7).

```
Ridge Regression Statistics 1:

         Variance     Bias^2       MSE         F     R2 adj-R2  rsigma2
K=0     8626.8525     0.0000  8626.853 1391.3951 0.9854 0.9848  40.5350
K=0.001 7597.0579   141.9010  7738.959 1390.7822 0.9846 0.9840  40.5528
K=0.002 6770.2743   510.1914  7280.466 1389.1517 0.9839 0.9832  40.6004
K=0.003 6093.2526  1038.4927  7131.745 1386.7571 0.9832 0.9825  40.6705
K=0.004 5529.6500  1679.5794  7209.229 1383.7914 0.9826 0.9819  40.7577
K=0.005 5053.8526  2399.2746  7453.127 1380.4063 0.9819 0.9812  40.8576
K=0.006 4647.3440  3172.5233  7819.867 1376.7147 0.9813 0.9806  40.9672
K=0.007 4296.4114  3980.7806  8277.192 1372.8044 0.9808 0.9800  41.0839
K=0.008 3990.6959  4810.2259  8800.922 1368.7428 0.9802 0.9794  41.2058
K=0.009 3722.2367  5650.5154  9372.752 1364.5802 0.9796 0.9788  41.3315
K=0.01  3484.8110  6493.8924  9978.703 1360.3576 0.9791 0.9783  41.4598
K=0.011 3273.4960  7334.5447 10608.041 1356.1045 0.9786 0.9777  41.5898
K=0.012 3084.3390  8168.1326 11252.472 1351.8459 0.9780 0.9772  41.7208
K=0.013 2914.1386  8991.4389 11905.577 1347.5998 0.9775 0.9766  41.8523
K=0.014 2760.2772  9802.1063 12562.383 1343.3779 0.9770 0.9761  41.9838
K=0.015 2620.5856 10598.4394 13219.025 1339.1917 0.9765 0.9756  42.1151
K=0.016 2493.2558 11379.2533 13872.509 1335.0498 0.9761 0.9751  42.2457
K=0.017 2376.7760 12143.7575 14520.533 1330.9557 0.9756 0.9746  42.3757
K=0.018 2269.8610 12891.4668 15161.328 1326.9148 0.9751 0.9741  42.5047
K=0.019 2171.4208 13622.1315 15793.552 1322.9292 0.9746 0.9736  42.6328
K=0.02  2080.5180 14335.6835 16416.201 1319.0028 0.9742 0.9731  42.7597
K=0.021 1996.3521 15032.1940 17028.546 1315.1340 0.9737 0.9726  42.8855
K=0.022 1918.2283 15711.8402 17630.069 1311.3225 0.9733 0.9722  43.0101
K=0.023 1845.5362 16374.8796 18220.416 1307.5719 0.9728 0.9717  43.1335
```

Figure 6. MSE, adj-*R*-square and other statistics produced by different *k* values for TL estimation

```
Ridge Regression Statistics 1:

           Variance     Bias^2       MSE        F     R2 adj-R2  rsigma2
K=0     127826.4729     0.0000 127826.47 168.5350 0.8908 0.8864 600.6178
K=0.001 112516.8563   282.9445 112799.80 168.5370 0.8897 0.8852 600.6110
K=0.002 100153.0506  1019.8292 101172.88 168.5388 0.8886 0.8841 600.6044
K=0.003  89983.5825  2082.7516  92066.33 168.5367 0.8876 0.8831 600.6119
K=0.004  81489.7399  3381.6844  84871.42 168.5286 0.8867 0.8821 600.6409
K=0.005  74302.2218  4851.7265  79153.95 168.5140 0.8858 0.8812 600.6930
K=0.006  68151.6381  6445.2066  74596.84 168.4925 0.8849 0.8802 600.7694
K=0.007  62836.8661  8126.5912  70963.46 168.4644 0.8840 0.8793 600.8696
K=0.008  58204.8844  9869.0876  68073.97 168.4299 0.8832 0.8785 600.9927
K=0.009  54137.3490 11652.3121  65789.66 168.3892 0.8824 0.8776 601.1380
K=0.01   50541.2370 13460.6529  64001.89 168.3428 0.8816 0.8768 601.3036
K=0.011  47342.5983 15282.0960  62624.69 168.2911 0.8808 0.8760 601.4886
K=0.012  44481.8079 17107.3689  61589.18 168.2345 0.8801 0.8752 601.6908
K=0.013  41910.3969 18929.3079  60839.70 168.1735 0.8793 0.8744 601.9091
K=0.014  39588.6056 20742.3832  60330.99 168.1082 0.8786 0.8737 602.1429
K=0.015  37483.4033 22542.3398  60025.74 168.0391 0.8779 0.8730 602.3906
K=0.016  35567.2001 24325.9223  59893.12 167.9666 0.8772 0.8722 602.6507
K=0.017  33816.8752 26090.6628  59907.54 167.8907 0.8765 0.8715 602.9231
K=0.018  32212.7526 27834.7164  60047.47 167.8119 0.8758 0.8708 603.2062
K=0.019  30738.1179 29556.7318  60294.85 167.7303 0.8752 0.8701 603.4995
K=0.02   29378.5915 31255.7502  60634.34 167.6465 0.8745 0.8695 603.8013
K=0.021  28121.8764 32931.1248  61053.00 167.5603 0.8739 0.8688 604.1119
K=0.022  26957.2962 34582.4571  61539.75 167.4718 0.8732 0.8681 604.4311
K=0.023  25875.4768 36209.5461  62085.02 167.3817 0.8726 0.8675 604.7564
```

Figure 7**.** MSE, adj-*R*-square and other statistics produced by different *k* values for TW estimation

3.3.3. Searching of the best *k* by acceptable VIF

Figure 8 gives the VIF values of each estimator variable for different *k*-ridge parameter values for TL and TW modelling. In contrast to the k values increased, VIF values were values decreased. The first *k* value where the VIF values of the predictor variables were all calculated below 10 was found to be 0.021.

3.3.4. Final decision on the optimum *k* parameter

So far, to decide the optimum k parameter, we have used the graphical method, calculating MSE, VIF and *R*-squared values. In graphical methods, we considered Ridge Trace Plot and VIF Trace Plot. In the Ridge Trace Plot, the *k* value that produces the lowest MSE is also given on the graph. For TL, *k*=0.003 and MSE=7131.745, for TW, *k*=0.016 and MSE=59893.122 values were produced. In VIF Trace Plot, the *k* value that produces the minimum GCV value is also given on the graph. While the *k* value that produced the lowest GCV for TL (0.325) was 0.003, the *k* value that produced the minimum GCV for TW (4.757) was found to be 0.018.

```
> vif(mod)
              PR       PD       HH       BH       HL      ED
k=0     60.84427 36.21742 46.01677 14.97258 50.27037 4.50356
k=0.001 51.68972 32.35179 41.03699 14.06270 43.77091 4.42522
k=0.002 44.53242 29.15939 36.86107 13.28556 38.56067 4.35466
k=0.003 38.82066 26.47946 33.31653 12.61041 34.30272 4.29006
k=0.004 34.18275 24.19909 30.27697 12.01567 30.76668 4.23016
k=0.005 30.36048 22.23654 27.64743 11.48566 27.79003 4.17403
k=0.006 27.16968 20.53107 25.35509 11.00873 25.25501 4.12101
k=0.007 24.47595 19.03654 23.34313 10.57601 23.07430 4.07060
k=0.008 22.17925 17.71724 21.56655 10.18061 21.18183 4.02241
k=0.009 20.20378 16.54507 19.98926  9.81710 19.52677 3.97614
k=0.01  18.49121 15.49760 18.58198  9.48111 18.06934 3.93154
k=0.011 16.99604 14.55672 17.32073  9.16911 16.77804 3.88843
k=0.012 15.68230 13.70761 16.18569  8.87818 15.62758 3.84665
k=0.013 14.52127 12.93807 15.16034  8.60590 14.59747 3.80607
k=0.014 13.48974 12.23792 14.23081  8.35026 13.67089 3.76658
k=0.015 12.56881 11.59862 13.38540  8.10952 12.83397 3.72809
k=0.016 11.74294 11.01296 12.61416  7.88222 12.07513 3.69053
k=0.017 10.99925 10.47480 11.90857  7.66711 11.38466 3.65382
k=0.018 10.32701  9.97889 11.26132  7.46308 10.75433 3.61792
k=0.019  9.71720  9.52069 10.66610  7.26919 10.17718 3.58276
k=0.02   9.16218  9.09630 10.11745  7.08461  9.64721 3.54832
k=0.021  8.65547  8.70230  9.61059  6.90859  9.15929 3.51454
k=0.022  8.19152  8.33573  9.14135  6.74049  8.70896 3.48140
k=0.023  7.76558  7.99398  8.70608  6.57973  8.29238 3.44886
k=0.024  7.37353  7.67474  8.30156  6.42578  7.90617 3.41690
k=0.025  7.01182  7.37600  7.92495  6.27819  7.54738 3.38550
k=0.026  6.67733  7.09595  7.57371  6.13654  7.21341 3.35462
k=0.027  6.36735  6.83299  7.24561  6.00043  6.90199 3.32426
k=0.028  6.07951  6.58570  6.93865  5.86954  6.61108 3.29438
```

Figure 8. Effect of different *k*-ridge values on the VIF value of each estimator variable for the total length (TL) and total weight (TW). Since VIF values decreased significantly, no further evaluation on *k* values were presented in the figure

When we searched for the minimum MSE values for different *k* values one by one, it was seen that the *k* value that produced the minimum MSE value was the same as the Rigge Trace Plot for TL and TW predictions. However, the first *k* value where each predictor variable in the model met the condition of a VIF value less than 10 was found to be *k*=0.021 for both predictions. In this case, it will be necessary to choose between the values of *k*=0.003 and *k*=0.0214 for TL and the values of *k*=0.016, *k*=0.018 and *k*=0.0214 for TW.

Although the lowest MSE was obtained when *k*=0.003 was selected for TL (7131.745), the VIF value of no predictor variable, except for the ED predictor variable (4.29), did not fall below the acceptable threshold value (VIF<10). On the other hand, if *k*=0.021 is selected, the MSE will increase slightly (17028.546) and the adj-*R*-square value will decrease slightly (from 0.9825 to 0.9726), and the VIF values of the predictor variables will be reduced to acceptable limits (VIF<10) (Figure 9).

```
> vif(mod)
              PR       PD       HH       BH       HL      ED
k=0.003 38.82066 26.47946 33.31653 12.61041 34.30272 4.29006
k=0.021  8.65547  8.70230  9.61059  6.90859  9.15929 3.51454
> rstats1(mod)

Ridge Regression Statistics 1:

        Variance    Bias^2       MSE        F     R2 adj-R2 rsigma2       CN
K=0.003 6093.253  1038.493  7131.745 1386.757 0.9832 0.9825 40.6705 420.4031
K=0.021 1996.352 15032.194 17028.546 1315.134 0.9737 0.9726 42.8855 180.6727
```

Figure 9. VIF (VIF) values and some statistics (rstats1) for *k*={0.003, 0.021} values for TL estimation

If k=0.016 and k=0.018 are selected for TW, lower MSE (59893.12 and 60047.47) and higher adj-R-square (0.8722 and 0.8708) values can be obtained. However, when k=0.016, the VIF values of the PR, PD, HH and HL predictor variables are greater than the acceptable threshold value (VIF>10). On the other hand, if k=0.021 is selected, although the MSE values are slightly high (61053.00) and the adj-R-square (0.8688) value is slightly low, the VIF values of the predictor variables drop to an acceptable level (VIF<10) (Figure 10).

```
> vif(mod)
              PR       PD       HH      BH       HL      ED
k=0.016 11.74294 11.01296 12.61416 7.88222 12.07513 3.69053
k=0.018 10.32701  9.97889 11.26132 7.46308 10.75433 3.61792
k=0.021  8.65547  8.70230  9.61059 6.90859  9.15929 3.51454
> rstatsl(mod)

Ridge Regression Statistics 1:

        Variance   Bias^2      MSE        F     R2 adj-R2  rsigma2       CN
K=0.016 35567.20 24325.92 59893.12 167.9666 0.8772 0.8722 602.6507 214.5852
K=0.018 32212.75 27834.72 60047.47 167.8119 0.8758 0.8708 603.2062 199.5919
K=0.021 28121.88 32931.12 61053.00 167.5603 0.8739 0.8688 604.1119 180.6727
```

Figure 10. VIF (VIF) values and some statistics (rstatsl) for k={0.016, 0.018, 0.021} values for TW estimation

Considering all the above findings, larger k values have the potential to both bring the regression coefficients closer to zero and produce lower adj-R-squared values. In addition, lower k values have the potential to produce higher VIF values and some potentially important predictor variables to be insignificant. Therefore, it was decided to take the optimum ridge k value as 0.021 for TL and TW predictions.

3.4. Ridge regression analysis results

The ridge regression result for TL and TW estimates based on k=0.021 value were given in Figure 11 and Figure 12, respectively. When looking at the ridge regression results in relation to TL, except for the HH and ED variables, the coefficients for the other predictor variables were found to be significant (p<0.05). The effects of predictor variables on TL were determined as HL>PR>PD>BH, from largest to smallest. The regression equation was estimated as follows:

TL= -0.96 + 0.79PR + 0.70PD + 0.45BH + 1.67HL
(Adj-R-square= 0.9726)

```
Coefficients: for Ridge parameter K= 0.021
           Estimate Estimate (Sc) StdErr (Sc) t-value (Sc) Pr(>|t|)
Intercept   -0.9739   -28860.2432   2010.7854     -14.3527   <2e-16 ***
PR           0.7901      177.3161     19.2664       9.2034   <2e-16 ***
PD           0.6966      144.3384     19.3184       7.4715   <2e-16 ***
HH           0.3102       25.0948     20.3016       1.2361   0.2187
BH           0.4484       54.3799     17.2127       3.1593   0.0020 **
HL           1.6704      177.2640     19.8192       8.9441   <2e-16 ***
ED           0.5447        6.3969     12.2769       0.5210   0.6033
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Ridge Summary
      R2    adj-R2  DF ridge         F       AIC       BIC
  0.9737    0.9726    4.0019 1315.1340  491.5358 1135.7909
Ridge minimum MSE= 17028.55 at K= 0.021
P-value for F-test ( 4.0019 , 125.0206 ) = 4.210836e-101
```

Figure 11. Results of Ridge regression in relation to TL according to parameter k=0.021

According to the ridge regression results in relation to TW, except for PD and ED variables, regression coefficients for other predictor variables were found to be significant (p<0.05). The effects of predictor variables on TW were determined as BH>HL>HH>PR from largest to smallest. The regression equation for the relationship was estimated as follows:

TW= -143.99+0.76PR+2.43HH+2.91BH+1.95HL
(Adj-R-square= 0.868)

```
Coefficients: for Ridge parameter K= 0.021
           Estimate Estimate (Sc) StdErr (Sc) t-value (Sc) Pr(>|t|)
Intercept -143.9605   -31550.0963   7546.8960      -4.1805   0.0001 ***
PR           0.7636      171.3529     72.3109       2.3697   0.0193 *
PD          -0.3423      -70.9362     72.5063      -0.9783   0.3298
HH           2.4353      197.0011     76.1963       2.5854   0.0109 *
BH           2.9181      353.8624     64.6031       5.4775   <2e-16 ***
HL           1.9522      207.1725     74.3857       2.7851   0.0062 **
ED          -7.6056      -89.3220     46.0779      -1.9385   0.0548 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Ridge Summary
      R2    adj-R2  DF ridge        F      AIC       BIC
  0.8739    0.8688    4.0019 167.5603 835.4152 1479.6703
Ridge minimum MSE= 61053 at K= 0.021
P-value for F-test ( 4.0019 , 125.0206 ) = 3.102247e-49
```

Figure 12. Results of Ridge regression in relation to TW according to parameter k=0.021

4. Discussion

For over three decades, ridge regression has proven to be a valuable tool for applied statisticians and should be routinely investigated in the context of collinear multiple regression (McDonald, 2009). As a regularized regression method, ridge regression reduces the chances of overestimating coefficients, and makes predictive and explanatory models more informative and interpretable, especially for datasets with strong collinearity. The parameter estimation also has less variance than that generated by iterating variables and is therefore more stable (Niu et al., 2019). Ridge regression adjusts the least squares estimator by reducing the impact of multicollinearity on parameter estimates. The resulting equation consists of least squares and estimates that are not biased but are also more stable; This is an important stability for models to be used in prediction (Rikardsen & Johansen, 2003). Sahin et al (2018) tried to estimate the impact of some traits on the albumen index value that is one of the interior egg quality traits using the Least Squares (LS), Ridge Regression (RR) and Principal Components Regression (PCR). Due to the multicollinearity between independent variables, the standard errors and VIF values of the partial regression coefficients in the LS method were found to be quite high. They concluded that the use of RR and PCR analysis methods could be more accurate instead of the LS method under multicollinearity problem.

OLS (k=0 case) and RR (k=0.021) regression analysis results for TL and TW predictions are given together in Figure 13 and Figure 14. When looking at the OLS and RR regression results in Figure 13, it is seen that the regression coefficients are different and the importance levels of the BH and PD predictor variables in RR also increase. However, as expected in RR (Ölmez & Yağanoğlu, 2023), there was a slight decrease in the adj-R-square value (from 0.9848 to 0.97).

```
Coefficients: for Ridge parameter K= 0
           Estimate Estimate (Sc) StdErr (Sc) t-value (Sc) Pr(>|t|)
Intercept   -4.0265   -31703.5376   4563.6421      -6.9470   <2e-16 ***
PR           0.9487      212.8959     49.6621       4.2869   <2e-16 ***
PD           0.7481      155.0240     38.3154       4.0460   0.0001 ***
HH          -0.9275      -75.0313     43.1890      -1.7373   0.0848 .
BH           0.4487       54.4057     24.6356       2.2084   0.0291 *
HL           2.2365      237.3432     45.1410       5.2578   <2e-16 ***
ED           0.1902        2.2332     13.5112       0.1653   0.8690
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Ridge Summary
      R2     adj-R2   DF ridge          F        AIC        BIC
   0.98540    0.98480    5.99995 1391.39510  487.13857 1137.12311
Ridge minimum MSE= 8626.852 at K= 0
P-value for F-test ( 5.99995 , 124.0001 ) = 3.530574e-111
--------------------------------------------------------------------


Coefficients: for Ridge parameter K= 0.021
           Estimate Estimate (Sc) StdErr (Sc) t-value (Sc) Pr(>|t|)
Intercept   -0.9739   -28860.2432   2010.7854     -14.3527   <2e-16 ***
PR           0.7901      177.3161     19.2664       9.2034   <2e-16 ***
PD           0.6966      144.3384     19.3184       7.4715   <2e-16 ***
HH           0.3102       25.0948     20.3016       1.2361   0.2187
BH           0.4484       54.3799     17.2127       3.1593   0.0020 **
HL           1.6704      177.2640     19.8192       8.9441   <2e-16 ***
ED           0.5447        6.3969     12.2769       0.5210   0.6033
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Ridge Summary
      R2    adj-R2  DF ridge          F       AIC       BIC
   0.9737    0.9726    4.0019 1315.1340  491.5358 1135.7909
Ridge minimum MSE= 8626.852 at K= 0
P-value for F-test ( 4.0019 , 125.0206 ) = 4.210836e-101
```

Figure 13. OLS (k=0) and RR (k=0.021) regression analysis results for TL prediction

When OLS and RR regression results for TW are compared (Figure 14), it is seen that the regression coefficients of the variables have changed. It was determined that the significance level of the HL variable increased in RR, and although the regression coefficient of the ED variable was significant in OLS (p<0.05), it was found to be insignificant in RR (p>0.05). As in the TL estimation, the adj-R-square value decreases slightly when TW is modeled with RR (from 0.87220 to 0.8685).

```
Coefficients: for Ridge parameter K= 0
           Estimate Estimate (Sc) StdErr (Sc) t-value (Sc) Pr(>|t|)
Intercept -144.7147   -29911.3869  17566.9195      -1.7027   0.0911 .
PR           0.9802      219.9666    191.1652       1.1507   0.2521
PD          -1.0923     -226.3406    147.4884      -1.5346   0.1274
HH           3.0434      246.1986    166.2483       1.4809   0.1412
BH           3.0943      375.2203     94.8304       3.9568   0.0001 ***
HL           2.4827      263.4704    173.7621       1.5163   0.1320
ED          -9.2464     -108.5916     52.0088      -2.0879   0.0389 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Ridge Summary
      R2     adj-R2   DF ridge          F        AIC        BIC
   0.89080    0.88640    5.99995  168.53505  837.59178 1487.57632
Ridge minimum MSE= 61053 at K= 0.021
P-value for F-test ( 5.99995 , 124.0001 ) = 3.830558e-57
--------------------------------------------------------------------


Coefficients: for Ridge parameter K= 0.021
           Estimate Estimate (Sc) StdErr (Sc) t-value (Sc) Pr(>|t|)
Intercept -143.9605   -31550.0963   7546.8960      -4.1805   0.0001 ***
PR           0.7636      171.3529     72.3109       2.3697   0.0193 *
PD          -0.3423      -70.9362     72.5063      -0.9783   0.3298
HH           2.4353      197.0011     76.1963       2.5854   0.0109 *
BH           2.9181      353.8624     64.6031       5.4775   <2e-16 ***
HL           1.9522      207.1725     74.3857       2.7851   0.0062 **
ED          -7.6056      -89.3220     46.0779      -1.9385   0.0548 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Ridge Summary
      R2    adj-R2  DF ridge          F       AIC       BIC
   0.8739    0.8688    4.0019  167.5603  835.4152 1479.6703
Ridge minimum MSE= 61053 at K= 0.021
P-value for F-test ( 4.0019 , 125.0206 ) = 3.102247e-49
```

Figure 14. OLS (k=0) and RR (k=0.021) regression analysis results for TW prediction

Among some measure on collinearity, it was pointed that the VIF (individual measure) had a better performance on the fitted model on the morphology of *C. macropomum*. Also, the presence of collinearity indicates that a substantial part of the information in one or more of these covariates is redundant (Rivas Villegas et al., 2024). Therefore, the higher adj-R-square values in OLS compared to RR could be expected. As supported in our case, the relationships of both TL and TW with the predictor variables, a small decrease was observed in the adjusted R-square values obtained by ridge estimation compared to the adjusted R-square values obtained by the OLS estimation method. However, a significant decrease in VIF values outweighs this drawback, indicating that the models are more stable. In the study, estimates were made using all fish caught, without distinguishing between male and female individuals. In a purely ecological study, this situation can be considered a limiting situation in terms of the generalizability of the study results. However, this can be ignored since the current study was not an ecological study and the focus of the study was how ridge regression was used methodologically in evaluating the fish morphology-growth relationship.

5. Conclusion

Fish growth occurs in length and weight. Although it is known that there is a relationship between morphometric characteristics fish growth, modeling is limited to simple linear regression due to the high level of multicollinearity among predictive variables. One of the methods used in cases of multicollinearity is the ridge regression method. However, in this approach, determining the optimum ridge regression parameter (penalization term, k or lambda) can often be speculative and difficult. In the study, using multiple linear models for total length (TL) and total weight (TW) ($Y = \alpha + \beta_1 X_1 + ... + \beta_n X_n + \varepsilon$), the holistic effect of metric features on these two features was estimated using ridge regression. Measurements of a total of 126 *Capoeta damascina* individuals sampled from the Euphrates River (around Adıyaman province, Türkiye) were used as the data set. Predorsal length (PR), head length (HL), post dorsal length (PD), head height (HH) and eye diameter (ED) predictive metric features were considered. The ridge optimum k-parameter, which produces acceptable values (VIF<10) of the variance inflation factor (VIF) of the predictor variables, which are indicators of multicollinearity in the modeling, was determined as k=0.021 in both models. With the Ridge estimation method, in the relationship between total height (TL) and predictive metric traits ($TL = -0.96 + 0.79PR + 0.70PD + 0.45BH + 1.67HL$), except for ED and HH, other predictive variables were statistically significant (p<0.05) (Adj-R-square=0.97). Except for eye diameter and post dorsal length (PD), it was determined that the effects of other metric traits on total weight were significant (p<0.05), and the relationship between total length and important predictive features was estimated as $TW = -143.99 + 0.76PR + 2.43HH + 2.91BH + 1.95HL$

(Adj-*R*-square=0.87). The methodological approach and findings in this study may contribute to filling a gap in the literature regarding the relationships between fish growth and morphology. Additionally, it could enhance better growth predictions for different fish species, aiding sustainable fisheries management and the selective cultivation of desired fish traits in aquaculture by improving the understanding of morphometric features.

Acknowledgements

This research was supported by the Adıyaman University Scientific Research Projects (BAP) Coordination Unit, Project No. 2012/001, and was carried out with the legal consent of the Ministry of Food, Agriculture, and Livestock, General Directorate of Fisheries and Aquaculture (permission date 05.04.2012; permission number: 01515). Thank you to the relevant institutions**.**

Funding Information

This work was supported by Adıyaman University Scientific Research Projects (BAP) Coordination Unit, Project Number: 2012/001.

Conflict of Interest

The authors declare that they have no known competing financial or non-financial, professional, or personal conflicts that could have appeared to influence the work reported in this paper.

Ethical Statement

The paper is not currently being considered for publication elsewhere, and it reflects the authors' own research and analysis in a truthful and complete manner.

References

Adeboye, N. O., Fagoyinbo, I. S., & Olatayo, T. O. (2014). Estimation of the Effect of Multicollinearity on the Standard Error for Regression Coefficients. *Journal of Mathematics*, *10*(4), 16-20. https://doi.org/10.9790/5728-10411620

Albayrak, A. S. (2005). Çoklu doğrusal bağlanti halinde enküçük kareler tekniğinin alternatifi yanlı tahmin teknikleri ve bir uygulama. *Uluslararası Yönetim İktisat ve İşletme Dergisi*, *1*(1), 105-126. (in Turkish)

Alp, A., Akyüz, A., Kara, C., & Güneş, H. (2013). *Assessment of movements of Capoeta capoeta in Bulam Stream, Euphrates River, and efficiency of the fish passage* [Oral presentation]. International Symposium on Fisheries and Aquatic Sciences, Erzurum, Turkey.

Alp, A., Kara, C., Büyükçapar, H. M., & Bülbül, O. (2005). Age, Growth and Condition of *Capoeta capoeta angorae* Hanko 1924 from the Upper Water Systems of the River Ceyhan, Turkey. *Turkish Journal of Veterinary & Animal Sciences*, *29*(3), 665-676.

Arechavala-Lopez, P., Sanchez-Jerez, P., Bayle-Sempere, J. T., Sfakianakis, D. G., & Somarakis, S. (2012). Morphological Differences between Wild and Farmed Mediterranean Fish. *Hydrobiologia*, *679*, 217-231.

Asadollah S., Soofiani, N.M., Keivany, Y., & Hatami, R. (2017). Age and Growth of the Mesopotamian barb, *Capoeta damascina*, in Central Iran. *Iranian Journal of Fisheries Sciences*, *16*(2), 511-521. https://doi.org/10.22092/ijfs.2018.114675

Büyükuysal, M. Ç., & Öz, İ. İ. (2016). Çoklu Doğrusal Bağıntı Varlığında En Küçük Karelere Alternatif Yaklaşım: Ridge Regresyon. *Düzce Üniversitesi Sağlık Bilimleri Enstitüsü Dergisi*, *6*(2), 110-114. (in Turkish)

Clabaut, C., Bunje, P. M., Salzburger, W., & Meyer, A. (2007). Geometric Morphometric Analyses Provide Evidence for the Adaptive Character of the Tanganyikan Cichlid Fish Radiations. *Evolution*, *61*(3), 560-578. https://doi.org/10.1111/j.1558-5646.2007.00045.x

Daoud, J. I. (2017). Multicollinearity and Regression Analysis. *Journal of Physics: Conference Series*, *949*, 012009. https://doi.org/10.1088/1742-6596/949/1/012009

Forrester, J. B., & Kalivas, J. H. (2004). Ridge Regression Optimization Using a Harmonious Approach. *Journal of Chemometrics*, *18*(7-8), 372-384. https://doi.org/10.1002/cem.883

Gao, Y., Fu, Z., Yang, J., Yu, M., & Wang, W. (2024). Spatial–temporal Differentiation and Influencing Factors of Marine Fishery Carbon Emission Efficiency in China. *Environment, Development and Sustainability*, *26*(1), 453-478. https://doi.org/10.1007/s10668-022-02716-6

Golub, G. H., Heath, M., & Wahba, G. (1979). Generalized Cross-Validation as a Method for Choosing a Good Ridge Parameter. *Technometrics*, *21*(2), 215-223. https://doi.org/10.1080/00401706.1979.10489751

Karakaş, S. (2008). *Çoklu doğrusal bağlantı problemi ve yanlı regresyon tahmincileri* [Master's thesis, İstanbul Universty]. (in Turkish)

Katz, M. H. (2006). *Multivariable analysis, A practical guide for clinicians* (2nd ed.). Cambridge University Press.

Keppeler, F. W., Montaña, C. G., & Winemiller, K. O. (2020). The Relationship between Trophic Level and Body Size in Fishes Depends on Functional Traits. *Ecological Monographs*, *90*(4), e01415.

https://doi.org/10.1002/ecm.1415

Khalaf, G., & Shukur, G. (2005). Choosing Ridge Parameter for Regression Problems. *Communications in Statistics - Theory and Methods*, *34*(5), 1177-1182. https://doi.org/10.1081/STA-200056836

Kyritsi, S., & Moutopoulos, D. K. (2018). Mouth Size and Body Length Relations for Freshwater Fish Species. *Turkish Journal of Fisheries and Aquatic Sciences*, *18*(11), 1331-1332. https://doi.org/10.4194/1303-2712-v18_11_10

Lenhardt, M., Jarić, I., Cvijanović, G., Kolarević, J., Gačić, Z., Smederevac-Lalić, M., & Višnjić-Jeftić, Ž. (2012). Comparison of Morphological Characters between Wild and Cultured Sterlet (*Acipenser ruthenus* L.). *Slovenian Veterinary Research*, *49*(4), 177-84.

Liyandja, T. L., Armbruster, J. W., Poopola, M. O., & Stiassny, M. L. (2022). Evolutionary Convergence in Body Shape Obscures Taxonomic Diversity in Species of the African *Labeo forskalii* Group: Case Study of *L. parvus* Boulenger 1902 and *L. ogunensis* Boulenger 1910. *Journal of Fish Biology*, *101*(4), 898-913. https://doi.org/10.1111/jfb.15148

Mami, A. M., Abdalla, A., & Bin Ismaeil, E. H. (2021). Selecting the Optimal Value of Penalty Parameter K in Ridge Regression Estimators. *International Journal of Sciences: Basic and Applied Research*, *60*(5), 189-209.

Martins, B. O., Franco-Belussi, L., Siqueira, M. S., Fernandes, C. E., & Provete, D. B. (2021). The Evolution of Red Blood Cell Shape in Fishes. *Journal of Evolutionary Biology*, *34*(3), 537-548. https://doi.org/10.1111/jeb.13757

McDonald, G. C. (2009). Ridge Regression. Wiley *Interdisciplinary Reviews: Computational Statistics*, *1*(1), 93-100. https://doi.org/10.1002/wics.14

Mojekwu, T. O., & Anumudu, C. I. (2015). Advanced Techniques for Morphometric Analysis in Fish. *Journal of Aquaculture Research & Development*, *6*(8), 1-6. https://doi.org/10.4172/2155-9546.1000354

Nandi S., & Saikia S.K. (2015). Role of Morphometry in Determining the Feeding Success of Small Freshwater Fish Species: Multivariate Analysis of *Amblypharyngodon mola*, *Puntius ticto*, and *Esomus danricus*. *Archives of Polish Fisheries*, *23*, 79-90. https://doi.org/10.1515/aopf-2015-0009

Niu, S. Q., Haake, D. M., & Knouft, J. H. (2019). Gut Content Biomass of Stream Fishes: Associations with Streamflow Velocity, Population Density, and Prey Abundance at Two Spatial Scales. *Hydrobiologia*, *826*(1), 85-97. https://doi.org/10.1007/s10750-018-3719-4

Okamura, H., Yamashita, Y., & Ichinokawa, M. (2017). Ridge Virtual Population Analysis to Reduce the Instability of Fishing Mortalities in the Terminal Year. *ICES Journal of Marine Science*, *74*(9), 2427-2436. https://doi.org/10.1093/icesjms/fsx089

Oke, J., Akinkunmi, W. B., & Etebefia, S. O. (2019). Use of Correlation, Tolerance and Variance Inflation Factor for Multicollinearity Test. *GSJ*, *7*(5), 652-659.

Ölmez, P., & Yağanoğlu, A. M. (2023). Use of Ridge Regression Analysis in the Multicollinearity in Animal Science. *Journal of Biometry Studies*, *3*(2), 32-38. https://doi.org/10.61326/jofbs.v3i2.02

Orbach, D. N., Donovan, M., & Purchase, C. F. (2019). Sexually Selected Traits are Larger and More Variable in Male than Female Beach-spawning Capelin, *Mallotus villosus*. *Journal of Fish Biology*, *95*(6), 1385-1390. https://doi.org/10.1111/jfb.14145

Özkale, M. R., & Altuner, H. (2021). Bootstrap Selection of Ridge Regularization Parameter: A Comparative Study via a Simulation Study. *Communications in Statistics - Simulation and Computation*, *52*(8), 3820-3838. https://doi.org/10.1080/03610918.2021.1948574

Pino, M. M., Rivas, D. V., Carrera, C. O., Avila, N. C., Salazar, T. M., Díaz Merino, H., Gastañaduy, C. C., Ysa, R. S., Lozado, J. C., Santos, R. C., Gastañaduy, D. C., & Palacios, L. V. (2021). Lasso Model for Morphological Covariation Patterns between *Colossoma macropomum* and *Piaractus orinoquensis* × *Colossoma macropomum* Hybrid. *Journal of Biochemical Technology*, *12*(1), 45-48. https://doi.org/10.51847/pFILL2lk3G

Revelle, W., & Revelle, M. W. (2015). Package 'psych'. *The comprehensive R archive network*, 337(338).

Rikardsen, A. H., & Johansen, M. (2003). A Morphometric Method for Estimation of Total Lipid Level in Live Arctic Charr: A Case Study of Its Application on Wild Fish. *Journal of Fish Biology*, *62*(3), 724-734. https://doi.org/10.1046/j.1095-8649.2003.00062.x

Rivas Villegas, D., Rivero Alzamora, C., Carrera Osorio, C., Calderón Ramírez, L., Rojas Correa, L., Cáceres, N., Sánchez Palacios, J., Del Carpio Franco, C., Vásquez Grados, M., Salinas Cruz, L., Rojas Ponce, K., Figueroa Rodríguez, J. J., & Quipas Bellizza, M. M. (2024). Comparison of Collinearity Indices for Linear Models in Agricultural Trials. *OnLine Journal of Biological Sciences*, *24*(2), 195-207. https://doi.org/10.3844/ojbsci.2024.195.207

Sahin, M., Yavuz, E., & Uckardes, F. (2018). Multicollinearity Problem and Bias Estimates in Japanese Quail. *Pakistan Journal of Zoology*, *50*(2), 757-761. http://dx.doi.org/10.17582/journal.pjz/2018.50.2.757.761

Thompson, C. G., Kim, R. S., Aloe, A. M., & Becker, B. J. (2017). Extracting the Variance Inflation Factor and other Multicollinearity Diagnostics from Typical Regression Results. *Basic and Applied Social Psychology*, *39*(2), 81-90. https://doi.org/10.1080/01973533.2016.1277529

Turan, C. (1999). A Note on the Examination of Morphometric Differentiation among Fish Populations: The Truss System. *Turkish Journal of Zoology*, *23*(3), 259-264.

Ullah, M. I., Aslam, M., & Altaf, S. (2018). lmridge: A Comprehensive R Package for Ridge Regression. *The R Journal*, *10*(2), 326-346.

Williams, G. J. (2011). *Data mining with Rattle and R: The art of excavating data for knowledge discovery*. Springer.

Winans, G. A. (1987). *Using morphometric and meristic characters for identifying stocks of fish* [Oral presentation]. The Slock Identification Workshop, Florida, USA.

Yazıcı, R. & Yazıcıoğlu, O. (2020). Some Morphometric Relationships of Wels Catfish (*Silurus glanis* L., 1758) Inhabiting Sıddıklı Dam Lake (Kırşehir, Turkey). *Journal of Anatolian Environmental and Animal Sciences*, *5*(2), 199-204. https://doi.org/10.35229/jaes.718647

RESEARCH ARTICLE
Journal of Biometry Studies (2024) 4(1): 42-48
DOI: 10.61326/jofbs.v4i1.04
www.biometrystudies.com

Journal of Biometry Studies
PRENSİP
www.prensip.gen.tr

Development of double haploid lines using anther culture method with different F2 combinations

Berrin DUMLU[1], Umran KUCUKOZDEMIR[1], Halit KARAGOZ[1,*]

[1]East Anatolia Agricultural Research Institute, Erzurum/TÜRKİYE

[*]Corresponding author: halit.karagoz@tarimorman.gov.tr
Received: 30/05/2024, Accepted: 21/06/2024

Abstract

Breeders are turning to rapid breeding methods in order to respond to the increasing population for the breeding of the wheat plant, which has a very important place in human nutrition. One of the biotechnological methods to shorten wheat breeding, which takes 15-20 years with classical breeding methods, is double haploid. The aim of this study is to obtain double haploid lines by taking wheat genotypes that have reached the F2 generation after hybridization into anther culture and thus shorten the breeding period. When looking at the response to anther culture of the 15 different F2 combinations used in the study, F2-7 and F2-10 genotypes stood out as the most successful genotypes. It was observed that 4 different groups were formed among 15 genotypes according to the success rate. It was concluded that the success of anther culture varies depending on the genotypes used, the nutrient media and the suitability of laboratory conditions.

Keywords: Wheat, Double haploid, Speed breeding, F2 combination

Please cite this article as follows:

Dumlu, B., Kucukozdemir, U., & Karagoz, H. (2024). Development of Double Haploid Lines Using Anther Culture Method with Different F2 Combinations. *Journal of Biometry Studies*, *4*(1), 42-48. https://doi.org/10.61326/jofbs.v4i1.04

1. Introduction

Wheat, which meets a significant portion of people's daily protein and calorie needs, is the most cultivated plant species in the world and is the most important species that meets the food needs in Turkey. (Kucukozdemir et al., 2020; Kucukozdemir et al., 2021; Kucukozdemir et al., 2023). As of 2022, the wheat cultivation area in our country is approximately 6.6 million hectares, the production amount is 19.7 million tons and the grain yield per unit area is around 298 kg/da (TUIK, 2022; Dumlu, 2023). Wheat yield was 254 kg/da in 2009 (TUIK, 2010). This increase is directly proportional to the registration of high-yield varieties and their introduction into production. As in other plants, the "double haplodization" method, which is the method of obtaining 100% homozygous pure lines in the shortest time by doubling the chromosome numbers of haploid plants, is very important in wheat breeding (Keles et al., 2015). Obtaining pure lines in the classical breeding method requires long processes. Pure lines can be obtained after 10-12 years in foreign pollinated plants and 6-7 years in self-pollinated plants. This period can be reduced to 6-12 months with haploidy methods (Salantur et al., 2011). However, one of the most important bottlenecks in the double haploid method is genotype selection and the unknown success rate of these genotypes. This study was conducted using 15 different F2 genotypes to determine the suitability of these genotypes for the double haploid method and to develop genotypes with superior properties using the double haploid method.

2. Material and Methods

In the study, 15 different hybrid combinations were created and cultured at the F2 stage. Hybrid information of these combinations is given in Table 1.

Table 1. Hybrid information of the F2 genotypes used in the study

F2-1	Unknown/Alparslan
F2-2	Alacris /Alparslan
F2-3	Genessi/Alparslan
F2-4	Savalan/Grk//Pyn/Bau/3/Alparslan
F2-5	Müfitbey/Alparslan
F2-6	Unknown/Plk70//Frtl/3/Alparslan
F2-7	Ayyıldız /Alparslan
F2-8	Esperia/Alparslan
F2-9	Doğu 88/Alparslan
F2-10	Palandöken 97/Alparslan
F2-11	Bezostaya/Alparslan
F2-12	Unknown/Alparslan
F2-13	Unknown/Plk70//Frtl/3/Alparslan*
F2-14	Yugtına/Kauz/3/Agrı/Bjy//Vee/4/Alparslan*
F2-15	F130-L-1-12/5/Lov26//Lfn/Sdy(Es84-24)/3/Serı/4/Serı/6/F6038w12-1/7/Alparslan*

In the greenhouse, from each F2 combination; The pollen found inside the anthers on the spikes of wheat genotypes was taken from 30 spikes in the early-mid univalent (single-nucleated) period (by examining the pollen under binocular). The spikes taken from the greenhouse were placed in a erlenmeyer flask containing water and covered with a nylon bag. These spikes were kept at 4 °C for 14-15 days. After the preliminary cold application, the stems and leaves of the spikes were cleaned, then they were placed in an erlenenmayer containing 1/1 sterile water and 0.5% sodium hypochlorite and shaken for 20 minutes in a way that would not cause physical damage to the spikes. After these procedures, the spikes were washed 3-4 times with sterile water under a sterile cabinet. After the sterilization process, the upper and lower spikelets on the spikes were removed. 90 anthers from the middle spikelets were taken with sterile forceps. Anthers were transferred to previously prepared petri dishes containing MN6 liquid medium. In this study, MN6 liquid medium was used for somatic embryo formation (Chu & Hill, 1988). 90 anthers were placed in each petri dish in 4 replicates. To prevent contamination in the Petri dishes, the dishes were covered with paraffin and left in a dark incubator at 29 °C for 40 days. The formed haploid embryos were transferred to 190-II (Xingzhi & Han, 1984) medium for plant regeneration. Developing plantlets were transferred to test tubes containing 190-II nutrient medium and rooting was ensured. The media and chemical components used in the experiment are given in Table 2. Rooted haploid plantlets were transferred to pots. The roots of the plantlets that developed roots in the pot were washed under water to remove soil. Roots were treated with 0.5% Colchicine for 1 hour.

Table 2. Media and chemical components used in the study

MN6		190-II Cu			
KNO3	1150 mg	KNO_3	100	Pyridoxine HCl	0,5
/NH4/2SO4	100 mg	$/NH_4/_2SO_4$	200	Nicotinik acid	0,5
Ca/ NO3/2 x 4 H2O	100 mg	Ca/ $NO_3/2$x 4 H_2O	100	Meso-inositol	100
MgSO4 x 7 H2O	125 mg	KH_2PO_4	300	Sakkarose	30 g
KH2PO4	200 mg	$MgSO_4$ x 7 H_2O	200	NAA	0,5
KCl	35 mg	KCl	40	Kinetin	0,5
Fe-Na-EDTA	5 ml	Fe-Na-EDTA	20	$CuSO_4$ x 5 H_2O	0,5
Thiamin-HCl	1 ml	$MnSO_4$ x 4 H_2O	8	Gelrite	3 g
Maltose	80 g	$ZnSO_4$ x 7 H_2O	3	pH	5,7
2,4-D	1,5 mg	H_3BO_3	3		
Kinetin	0,5 mg	Kl	0,5		
Ficoll	100 g	Glycine	2		
pH	5,8	Thiamin-HCl	1		

Parameters examined in the study: *Number of calluses (NC)* refers to the number of calluses formed from anthers taken from each genotype in the study. *Number of green plants (NGP)* refers to the number of green plants obtained from calluses. *Callus formation rate from anther (CFA)* is the percentage expression of the callus consisting of 90 anthers taken per petri dish. *Double haploid index (DHI)* is the percentage expression of green plantlets obtained from calluses. *Success rate (SR)* is the percentage expression of the number of natural and colchicine-grown plants from callus. (Salantur et al., 2011).

The study was designed according to the randomized block trial design with 4 replications. The results of the study were made using the Jump 17 program for variation analysis and were subjected to the LSD multiple comparison test. Heatmapper graphics were made with the heatmapper online program, PCA and Venn graphics were made with the tbtools program, and Dendogram graphics were made with the Srplot online program.

3. Results and Discussion

According to the results of the variation analysis, it was determined that there were statistically significant (p<0.01) differences between the genotypes in the NC, NGP, CFA, DHI and SR parameters (Table 3). In terms of NC, the F2-10 (15.5) genotype received the highest value, while the lowest NC value was determined in the F2-1 (4.5) genotype. The average NC value of the genotypes was 12.03. When we look at the NGP numbers, the highest NGP number was determined in the F2-10 genotype and the lowest NGP number was determined in the F2-12 and F2-13 genotypes. The average NGP numbers of the genotypes were 1.65. While the average CFA of the genotypes was 13.37%, the highest CFA was in the F2-10 (17.22%) genotype and the lowest CFA was in the F2-1 (5%) genotype. The highest DHI value was measured in F2-5 (50.63) and the lowest DHI value was measured in F2-12 (0) and F2-13 (0) genotypes. The average DHI value of the genotypes was 14.37. The highest SR value in the study was determined in the F2-5 (6.67) genotype, and the lowest SR value in the F2-12 (0) and F2-13 (0) genotypes, while the average SR value of the genotypes was 1.83 (Table 3).

In the Heatmapper chart, the colors of the genotypes gradually increase from red to green, and as they approach black, they get closer to the average value. Accordingly, in terms of all applications, F2-10, F2-7 and F2-8 genotypes stood out with higher values than the average. It was observed that while F2-3, F2-14, F2-13, F2-12, F2-15 and F2-11 genotypes had above average values in terms of NC and CFA, they had below average values in terms of NGP, SR and DHI parameters. F2-5 genotype had values close to the average in terms of NC and CFA parameters and was the genotype with the highest values in terms of NGP, SR and DHI (Figure 1).

According to PCA analysis, it was seen that the genotypes were divided into 4 different groups. F2-5 genotype in Group 1, F2-2, F2-3, F2-6, F2-7, F2-8, F2-10 and F2-14 genotypes in Group 2, F2-1, F2-4 in Group 3, and F2-9 genotypes, and F2-11, F2-12, F2-13 and F2-15 genotypes were included in Group 4 (Figure 2).

In the dendogram analysis, similar to the PCA analysis, it was seen that the genotypes were divided into 4 different groups. F2-5 genotype in Group 1, F2-2, F2-3, F2-6, F2-7, F2-8, F2-10 and F2-14 genotypes in Group 2, F2-1, F2-4 in Group 3, and F2-9 genotypes, and F2-11, F2-12, F2-13 and F2-15 genotypes were included in Group 4 (Figure 3).

Venn chart is a type of graph in which the common intersection set of genotypes that stand out in terms of applications is expressed. According to the Venn graph in this study, it is seen that there are two genotypes that stand out in terms of all parameters. These genotypes are F2-7 and F2-10 genotypes (Figure 4).

Table 3. Parameters examined in the study

	NC*	NGP*	CFA*	DHI*	SR*
F2-1	4.50^{f}	0.50^{ef}	5.00^{f}	15.63^{b-e}	0.56^{ef}
F2-2	11.50^{c-e}	1.50^{d-e}	12.78^{c-e}	13.13^{b-e}	1.67^{d-e}
F2-3	14.00^{a-c}	2.00^{b-d}	15.56^{a-c}	13.95^{b-e}	2.22^{b-d}
F2-4	8.50^{de}	0.50^{ef}	9.44^{de}	5.00^{d-f}	0.56^{ef}
F2-5	12.00^{bc}	6.00^{a}	13.33^{bc}	50.63^{a}	6.67^{a}
F2-6	11.75^{b-d}	1.75^{cd}	13.06^{b-d}	$16,39^{b-d}$	1.94^{cd}
F2-7	14.25^{a-c}	3.00^{b}	15.83^{a-c}	21.40^{bc}	3.33^{b}
F2-8	12.50^{a-c}	2.75^{bc}	$13,89^{a-c}$	22.74^{b}	3.06^{bc}
F2-9	8.25^{e}	1.50^{d-e}	9.17^{e}	19.58^{bc}	1.67^{d-e}
F2-10	15.50^{a}	3.00^{b}	17.22^{a}	19.81^{bc}	3.33^{b}
F2-11	13.25^{a-c}	0.50^{ef}	14.72^{a-c}	3.57^{ef}	0.56^{ef}
F2-12	15.00^{ab}	0.00^{f}	16.67^{ab}	0.00^{f}	0.00^{f}
F2-13	13.50^{a-c}	0.00^{f}	15.00^{a-c}	0.00^{f}	0.00^{f}
F2-14	$12,75^{a-c}$	1.25^{d-e}	14.17^{a-c}	9.83^{c-f}	1.39^{d-e}
F2-15	13.25^{a-c}	0.50^{ef}	14.72^{a-c}	3.71^{ef}	0.56^{ef}
Mean	12.03	1.65	13.37	14.37	1.83

*Means shown with different letters are statistically significantly different from each other according to the p<0.01 significance level in the LSD test.

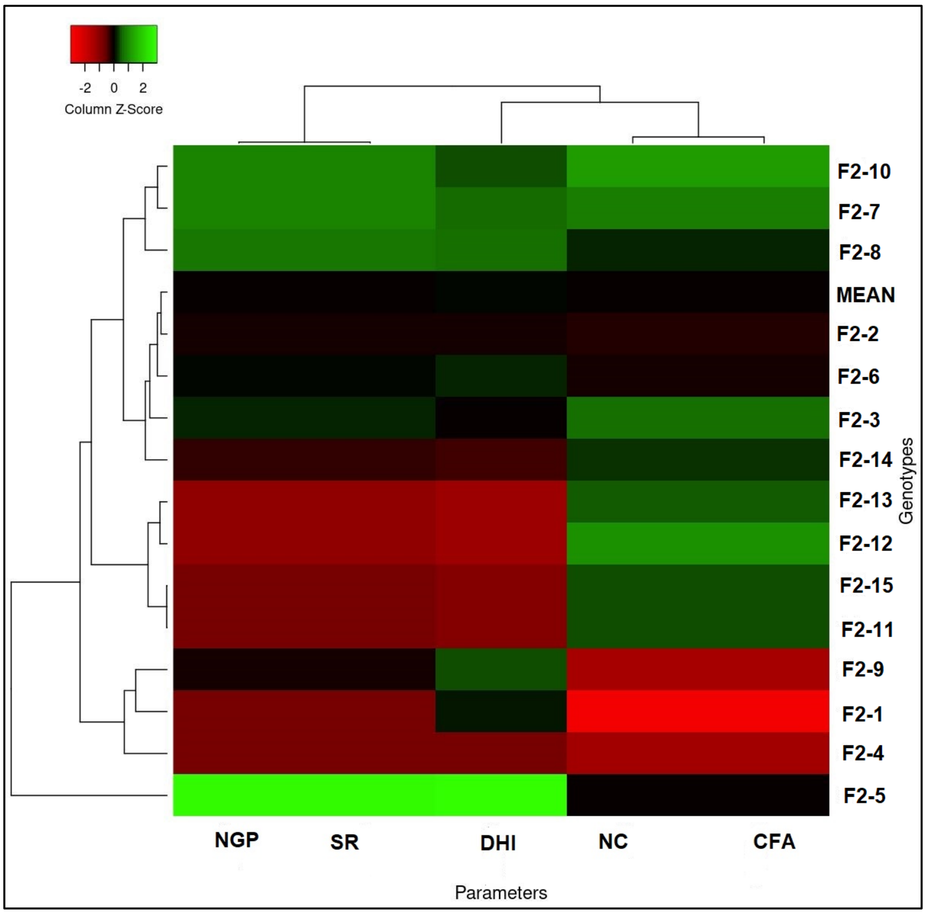

Figure 1. Heatmapper chart

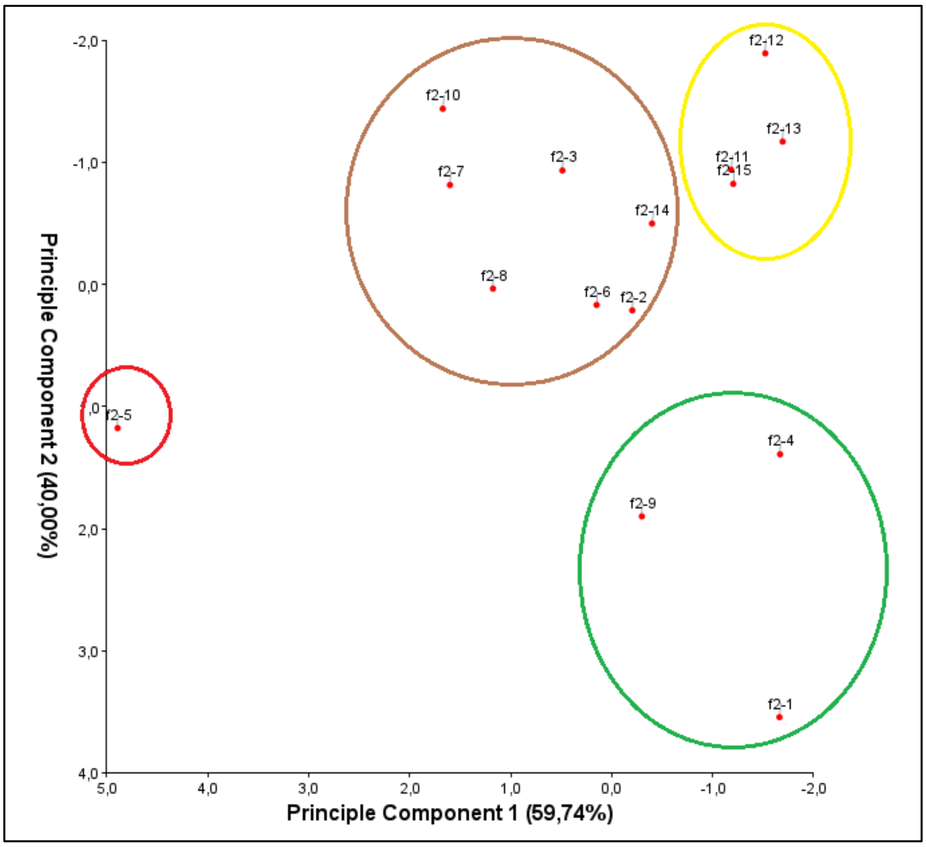

Figure 2. Principal component analysis

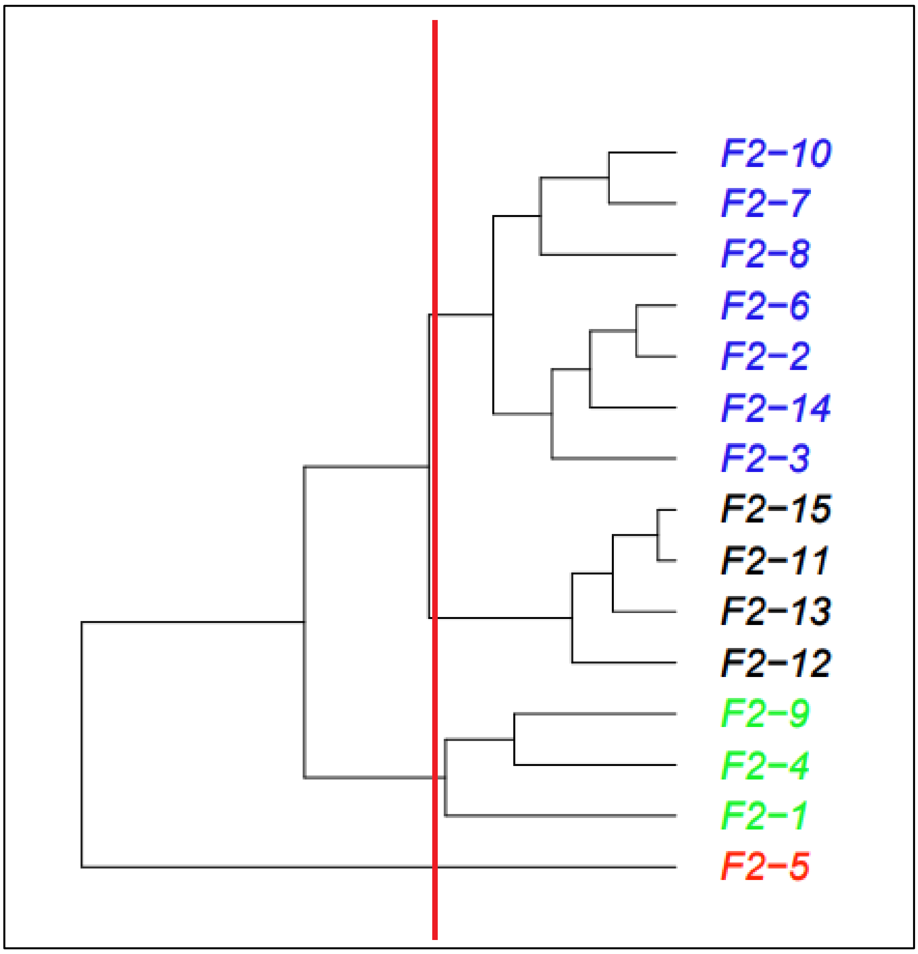

Figure 3. Dendrogram graph of 15 different genotypes

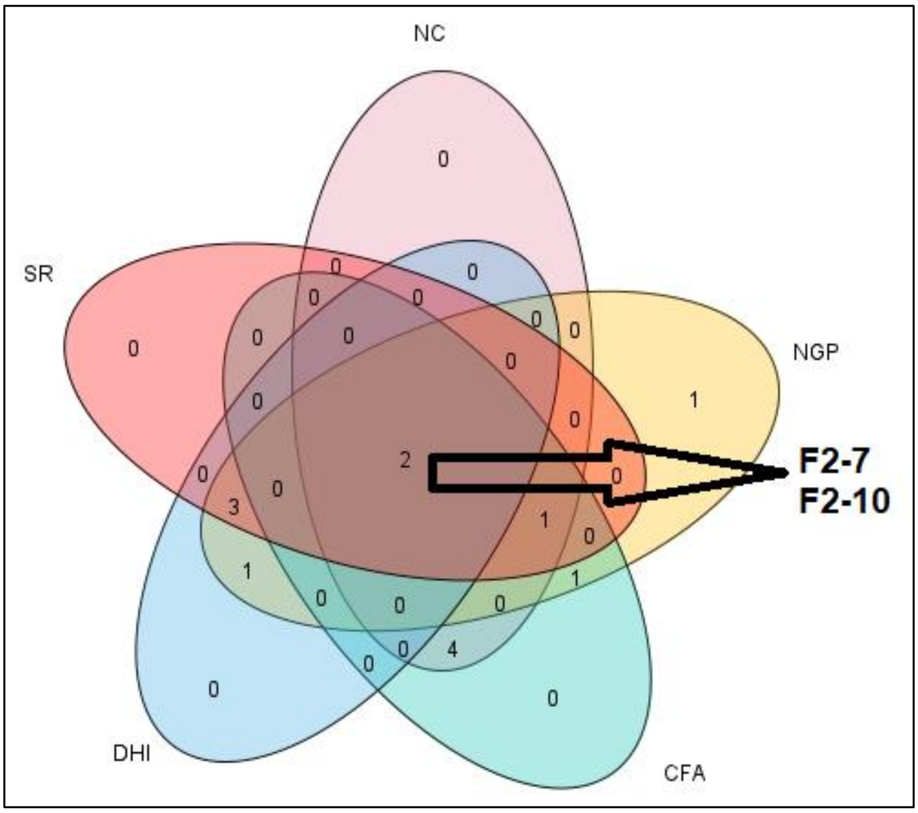

Figure 4. Venn chart of 15 different genotypes

4. Conclusion

In this study, the success rate obtained from anther culture varied depending on the genotypes. F2-7 and F2-10 genotypes in the study stood out as the most successful genotypes. Table 1, Figure 1 and 4). As a matter of fact, Szakacs et al. (1989) stated that despiegh the success of anther culture varies depending on the genotype; They reported that factors such as callus formation, plant regeneration, and the amount of double haploid plants success. In a study, it was stated that the rate of double haploid plants varies depending on not only the genotype but also the induction environment. (Orshinsky & Sadasivaiah, 1994). In another study, it was stated that pre-cold application increased the amount of naturally double haploids. (Pauk et al., 2003). Additionally, Bajaj (1990) reported that many factors such as the growing conditions of donor plants in anther culture, genotype, pre-applications to flower buds, structure and composition of the nutrient medium, and incubation conditions affect the success rate in anther culture. However, most researchers have stated that the biggest problem for the production of double haploid plants is still genotype dependency. (Barbanas, 2003; Inagaki, 2003; Zamani et al., 2003; Özü, 2006; Ahmet & Adak, 2007). In our study, the formation of 4 different groups among 15 genotypes according to the success rate is a good indicator of how success depends on the characteristics of the genotype. It is thought that hybrids made from Palandöken 97 and Ayyıldız varieties respond better to anther culture and can be used as parents. As a result, in order to achieve success in the double haploid method, which shortens the process in wheat breeding, which requires many years and labor, and saves time and labor for breeders, it is necessary to conduct studies on more genotypes and determine properties of this genotypes.

Acknowledgement

This article was produced from the Eastern Anatolian Wheat Breeding Research Project. We thank the General Directorate of Agricultural Research and Policy for supporting this project.

Conflict of interest

The authors declare that there is no conflict of interest.

Ethical Approval

For this type of study, formal consent is not required.

References

Ahmet, H., & Adak, M. S. (2007). Callus Induction and Plant Regeneration in Some Iraqi Common Wheat Varieties. *Journal of Agricultural Sciences*, *13*(3), 285-292.

Bajaj, Y. P. S. (1990). *Haploids in crop improvement I.* Springer-Verlag.

Barbanas B. (2003). Protocol for producing doubled haploid plants from anther culture of wheat (*Triticum aestivum* L.). In M. Maluszynski, K. J. Kasha, B. P. Forster, & I. Szarejko (Eds.), *Doubled Haploid Production in Crop Plants* (pp. 65-70). Springer.

Chu, C. C. & Hill, R. D. (1988). An Improved Anther Culture Method for Obtaining Higher Frequency of Pollen Embryoids in *Triticum aestivum* L. *Plant Science*, *55*(2), 175-181. https://doi.org/10.1016/0168-9452(88)90174-4

Dumlu, B. (2023). *Evaluation of Kırik wheat populations in terms of cold resistance, some agricultural features and quality* [PhD dissertation, Atatürk University]. (in Turkish)

Inagaki, M. N. (2003). Doubled haploid production in wheat through wide hybridization. In M. Maluszynski, K. J. Kasha, B. P. Forster, & I. Szarejko (Eds.), *Doubled Haploid Production in Crop Plants* (pp. 53-58). Springer.

Keles, D., Pinar, H., Ata, A., Taskin, H., Yildiz, S. & Buyukalaca, S. (2015) Effect of Pepper Types on Obtaining Spontaneous Doubled Haploid Plants via Anther Culture. *HortScience*, *50*(11), 1671-1676. https://doi.org/10.21273/HORTSCI.50.11.1671

Kucukozdemir, U., Dumlu, B., Yalcin, Z., & Karagoz, H. (2020). Determination of the Yieldy and Cold Hardiness Characteristics of Some Bread Wheat (*Triticum aestivum* L.) Genotypes in Aqueous Conditions. *Anatolian Journal of Biology*, *1*(1), 12-15.

Kucukozdemir, U., Dumlu, B., Karagoz, H., & Yilmaz, O. (2021). Determination of Yield, and Cold Hardiness of Some Triticale (xTriticosecale Wittmack) Genotypes in Eastern Anatolia Region. *Journal of Agricultural Production*, *2*(1), 26-31. https://doi.org/10.29329/agripro.2021.344.4

Kucukozdemir, U., Dumlu, B., Yilmaz, O., Turkoglu, A., Karagoz, H., & Kizilgeci, F. (2023). A Study On the Determination of Cold Resistance of Barley Genotypes with Different Cold Acclimatization Different Times and Growing Nature. *Fresenius Environmental Bulletin*, *32*(6), 2452-2459.

Orshinsky, B. R., & Sadasivaiah, R. S. (1994). Effects of Media on Embryoid Induction and Plant Regeneration from Cultured Anthers of Soft White Spring Wheats (*Triticum aestivum* L.). *Plant Science*, *102*(1), 99-107. https://doi.org/10.1016/0168-9452(94)90023-X

Özü, M. E. (2006). *Effect of growing conditions of donor plants and different incubation temperatures on* *haploid plant regeneration from anthers of bread wheat* [Master's thesis, Çukurova University]. (in Turkish)

Pauk, J., Mihaly, R. & Puolimatka, M. (2003). Protocol for wheat (*Triticum aestivum* L.) anther culture. In M. Maluszynski, K. J. Kasha, B. P. Forster, & I. Szarejko (Eds.), *Doubled Haploid Production in Crop Plants* (pp. 59-64). Springer.

Salantur, A., Yazar, S., Dönmez, E., & Akar, T. (2011). Determination of the Response of Winter Bread Wheat F2 Populations to Plant Regeneration in Anther Culture. *Journal of Field Crops Central Research Institute*, *20*(1), 15-21.

Szakacs, E., Kovacs, G., Pauk, J. & Barbanas, B. (1989). Substitution Analysis of Callus Induction and Plant Regeneration from Anter Culture in Wheat (*Triticum aestivum* L.). *Plant Cell Reports*, *7*, 218-222.

TUIK, (2010). Agricultural structure production, price, value. Turkish Statistical Institute, Ankara.

TUIK, (2022). Agricultural structure production, price, value. Turkish Statistical Institute, Ankara.

Xingzhi, W., & Han, H. (1984). The Effect of Potato II Medium for Triticale Anther Culture. *Plant Science Letters*, *36*(3), 237-239. https://doi.org/10.1016/0304-4211(84)90175-5

Zamani, I., Gouli-Vavdinoudi, E., Kovacs, G., Xynias, I., Roupakias, D., & Barbanas, B. (2003). Effect of Parental Genotypes and Colchicine Treatment on the Androgenic Response of Wheat F1 Hybrids. *Plant Breeding*, *122*, 314-317. https://doi.org/10.1046/j.1439-0523.2003.00866.x

RESEARCH ARTICLE
Journal of Biometry Studies (2024) 4(1): 49-55
DOI: 10.61326/jofbs.v4i1.05
www.biometrystudies.com

Journal of Biometry Studies
PRENSİP
www.prensip.gen.tr

Height−diameter relationships in Brutian pine stands in Kurşunlu Waterfall Nature Park

Dilek ŞEN[1], Fadime SAĞLAM[1,*]

[1]Kastamonu University, Faculty of Forestry, Department of Forest Engineering, Kastamonu/TÜRKİYE

[*]Corresponding author: fsaglam@kastamonu.edu.tr
Received: 29/05/2024, Accepted: 24/06/2024

Abstract

In the scope of this study, the aim was to develop height−diameter models for Brutian pine forests located within the boundaries of Kurşunlu Waterfall Nature Park in Aksu district of Antalya province. For this purpose, height−diameter data for 516 sample trees were randomly divided into model (422 trees) and control (94 trees) groups. Using height−diameter values obtained from 422 sample trees, parameter estimates were made for three models that are commonly used in the literature and successfully reflect height−diameter relationships. The best model was evaluated based on four success criteria: Coefficient of Determination (R^2), Root Mean Square Error (*RMSE*), Mean Absolute Error (*MSE*), and Akaike Information Criterion (*AIC*). The results of the evaluation indicated that the M2 (Meyer) model, which exhibited the lowest ranking value, was the most effective model. The observed and predicted height values of the trees in the independent data group were compared using the Student's *t*−test. The results of the test indicated that the observed and predicted height values were statistically indistinguishable (p>0.05). The height−diameter model developed as a result of the study will serve as a significant tool for biomass and carbon stock estimation.

Keywords: *Pinus brutia* Ten., Height-diameter models, Tree height, Diameter at breast height

Please cite this article as follows:
Şen, D., & Sağlam, F. (2024). Height−Diameter Relationships in Brutian Pine Stands in Kurşunlu Waterfall Nature Park. *Journal of Biometry Studies*, *4*(1), 49-55. https://doi.org/10.61326/jofbs.v4i1.05

1. Introduction

The diameter and height of trees represent fundamental measurements in forest inventory studies. The measurement of diameter and height of trees is employed as a fundamental variable in numerous applications within the field of forestry. These include the estimation of tree and stand volume, the development of growth models, the analysis of stand structural diversity and the estimation of biomass and carbon stock. The diameter at breast height of trees can be readily and precisely quantified to facilitate these estimations, whereas the measurement of tree height is more challenging and time−consuming. In this context, height−diameter models are of significant importance in the present era (Arabatzis & Burkhart, 1992; Huang et al., 2000; Sharma & Parton 2007; Meng et al., 2009; Diamantopoulou & Özçelik, 2012; Ercanlı, 2015).

A positive curvilinear relationship between height and diameter is observed in general. In order to model this relationship, both linear and non−linear regression models have been employed. Non−linear regression models are more flexible than linear regression models and provide ease of application to data. Consequently, they are employed with greater frequency (Larsen & Hann, 1987; Wang & Hann, 1988; Arabatzis & Burkhart, 1992). One of the models that form the basis for the growth and yield models required for ecosystem−based functional forest management planning used in Türkiye is the height−diameter model. Studies on these models are of great importance (Özçelik & Çapar, 2014).

When analyzing studies on height-diameter models in Türkiye, the following are observed Sönmez (2009), Mısır (2010), Özçelik et al. (2013), Özçelik and Çapar (2014), Ercanlı (2015), Çatal and Carus (2018), Özçelik et al. (2018), Ercanlı and Eyüboğlu (2019), Ercanlı (2020a, 2020b), Seki and Sakici (2022). Among these studies, the study conducted by Özçelik and Çapar (2014) was carried out in Antalya province.

It provides valuable predictions regarding stand structure, volume estimations for individual trees and stands, individual tree growth models, biomass and carbon estimations by modeling the relationships between diameter and height variables. *Pinus brutia*, the most widely distributed coniferous species in Türkiye, especially in Antalya province, offers numerous economic and ecological benefits. Consequently, height-diameter models developed for this species and region are of significant importance. In protected areas, the relationships between the diameter and height values of trees are of particular significance with regard to biomass and carbon estimation. The aim of this study is to develop height–diameter models for Brutian pine stands in the Kurşunlu Waterfall Nature Park.

2. Material and Methods

The General Directorate of Nature Conservation and National Parks is responsible for the operation of Nature Parks. Consequently, the natural structure of forest areas should be preserved. The Kurşunlu Waterfall, which has been designated a Nature Park, is protected in accordance with the provisions of the National Parks Law. The Kurşunlu Waterfall Nature Park is situated within the country coordinate system at latitudes 37° 00' 48" to 37° 00' 63" North and longitudes 30° 81' 00" to 30° 84' 05" East, in the Aksu district of Antalya province, Türkiye (Figure 1).

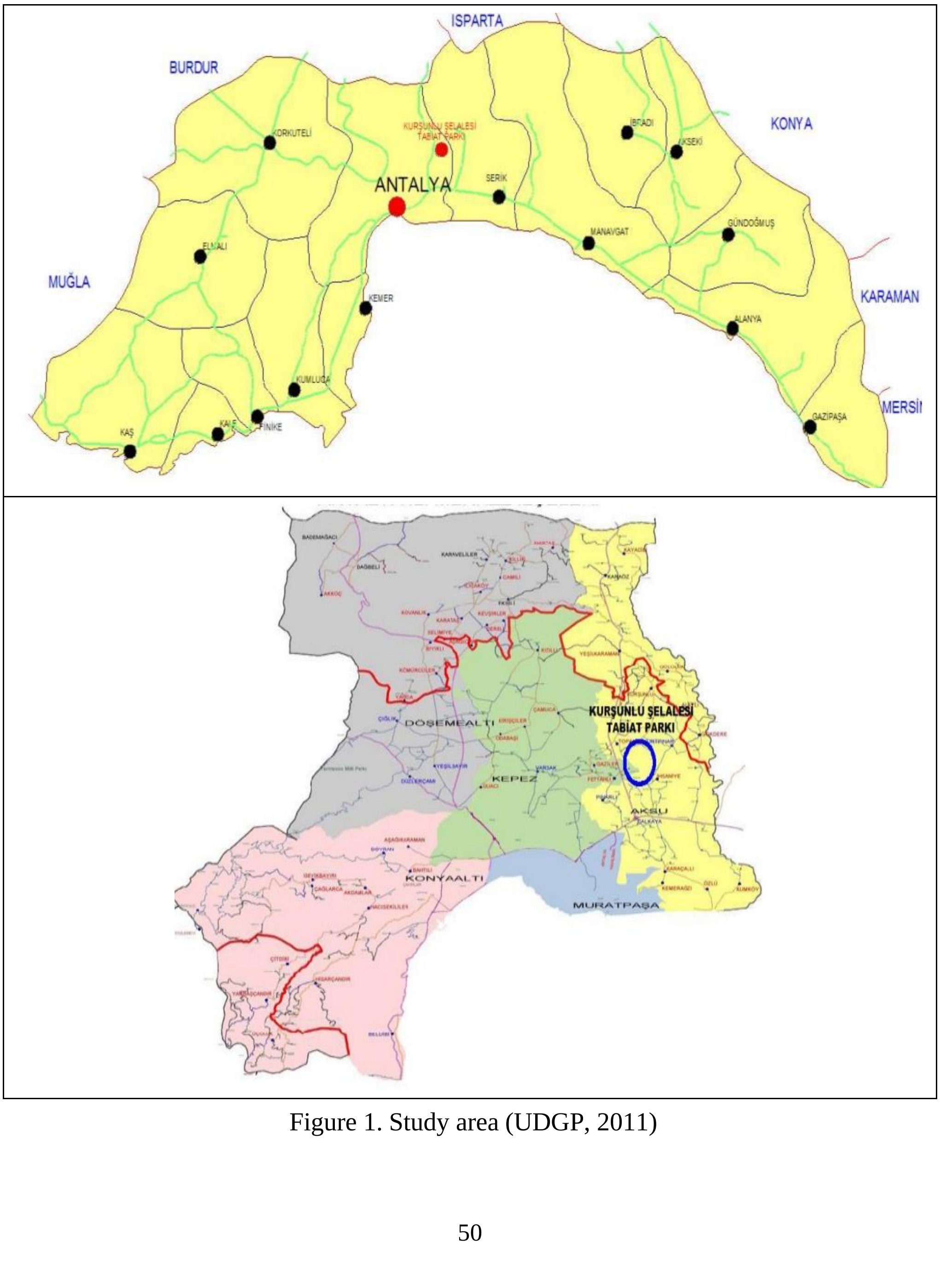

Figure 1. Study area (UDGP, 2011)

The Kurşunlu Waterfall Nature Park is located in a transitional zone between the Mediterranean climate and the terrestrial climate of Central Anatolia. The climate of the Park is generally characterized by hot and dry summers, and mild and rainy winters. In terms of temperature values in Türkiye, Antalya has a temperature above the average (UDGP, 2011).

The tree communities found in the Kurşunlu Waterfall Nature Park are mainly composed of Brutian pine (*Pinus brutia* Ten.), which is the dominant species in the Mediterranean region. In the scope of the study, height−diameter data obtained from 516 sample trees were used to develop height−diameter models for Brutian pine. These data were randomly divided into 2 groups (model and control) and the data from 422 sample trees were used to develop the models, while the data from 94 sample trees were used for the validation of the developed models (Figure 2). Non−linear regression analysis was used to estimate model parameters based on data from 422 trees. All parameters of the height−diameter models that were found to be statistically significant were subjected to a relative ranking based on various statistical criteria, and the best non−linear height−diameter model was determined.

In this study, three different basic non−linear height−diameter models were used, utilizing models used in the modelling of height−diameter relationships by various researchers. The height−diameter model structures for which parameter estimation was performed, together with the references used, are listed in Table 1.

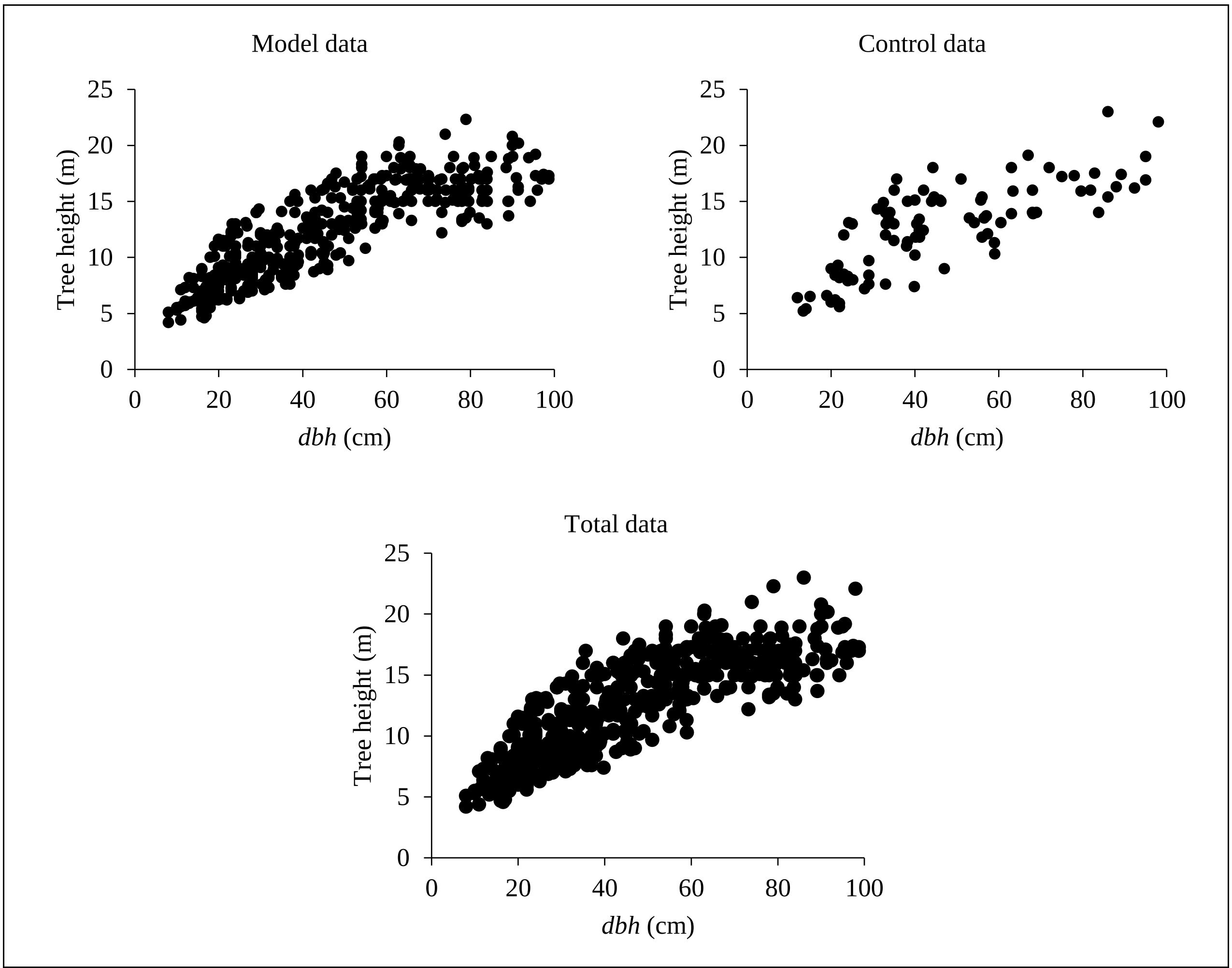

Figure 2. Height−diameter distributions for model, control and total data

Table 1. Models used to develop height−diameter models

Model	Model Form	References
M1	$h_{ij} = 1.3 + a * \left(1 - e^{(-b*dbh)^{c}}\right)$	Chapman-Richards (1959)
M2	$h_{ij} = 1.3 + a * \left(1 - e^{(-b*dbh)}\right)$	Meyer (1940)
M3	$h_{ij} = 1.3 + a * dbh^{b}$	Power function

h=tree height (m), *dbh* =diameter at breast height (cm) and *a*,*b*,*c*=model parameters

In this study, the height–diameter models were compared based on qualitative and quantitative evaluations. After parameter estimation of the models, the coefficient of determination (R^2) and various error statistics (Root Mean Square Error, *RMSE*; Mean Absolute Error, *MAE*; Akaike Information Criterion, *AIC*) were calculated for these models. After selecting the best height-diameter model according to statistical criteria and all parameters were significant, the model validity was tested. For this purpose, the observed height values of the trees of the independent data (control group) and the height values predicted from the best model were compared by Student's *t*–test.

The evaluation of the constructed models was based on four different statistical criteria. These statistical criteria are:

Coefficient of Determination:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(yi - \hat{y}i)^2}{\sum_{i=1}^{n}(yi - \bar{y}i)^2}$$

Root Mean Square Error:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(yi - \hat{y}i)^2}{n - p}}$$

Mean Absolute Error:

$$MAE = \frac{\sum_{i=1}^{n}|\hat{y}i - yi|}{n}$$

Akaike Information Criterion:

$$AIC = n\ ln(HKOK) + 2p$$

where yi, $\hat{y}i$, $\bar{y}i$ are the observed, predicted and average values of the dependent variable, respectively, n is the number of observations, and p is the number of model parameters. Subsequently, the error distributions were analyzed visually. Graphical representations were used to show deviations and systematic errors.

3. Results

Within the scope of the study, model parameters, various statistical criteria and performance rankings of 3 non–linear models, for which parameter estimations were made using data obtained from 422 sample trees, are presented in Table 2 and Table 3.

When examining Table 2, which shows the model parameters and statistical criteria, all parameters of the models were found to be significant at the $p<0.05$ level. Upon examination of Table 3 and Figure 3, it is observed that although the statistical criteria for the models have similar values, M2 stands out in the ranking.

To validate the model, the height predictions of the trees in the independent data subset (control groups; 94 trees) were made using the height–diameter model (M2). The observed heights of the trees in the independent data subset were compared with the predicted heights using Student's t–test, and the test results indicated that the observed and predicted heights were statistically indistinguishable ($p>0.05$).

The prediction and bias values for the developed height–diameter model were presented in Figure 4 for the model, the control and total data set. When Figure 4 is analyzed, it is observed that the bias values associated with predictions made by the height–diameter model exhibit a random distribution and do not have any trend. It is understood that the observed height values and the height values predicted by the M2 model are close to each other. As a result of these evaluations, it is ensured that the assumptions regarding regression analysis, such as bias being randomly distributed and having no trend, are met. Satisfying these assumptions is also an important result for the success of the model.

Table 2. Statistical criteria and parameter estimations for the models

Model	R^2	*RMSE*	*MAE*	*AIC*	Parameters		
					a	b	c
M1	0.765	2.015	1.612	301.686	19.692	0.019	0.940
M2	0.764	2.013	1.608	299.287	19.130	0.021	
M3	0.752	2.065	1.669	309.920	1.280	0.570	

Table 3. Success ranking of the models

Model	R^2	*RMSE*	*MAE*	*AIC*	Rank (R^2)	Rank (*RMSE*)	Rank (*MAE*)	Rank (*AIC*)	Total Rank
M1	0.765	2.015	1.612	301.686	1.0	2.0	2.0	2.0	7.0
M2	0.764	2.013	1.608	299.287	2.0	1.0	1.0	1.0	5.0
M3	0.752	2.065	1.669	309.920	3.0	3.0	3.0	3.0	12.0

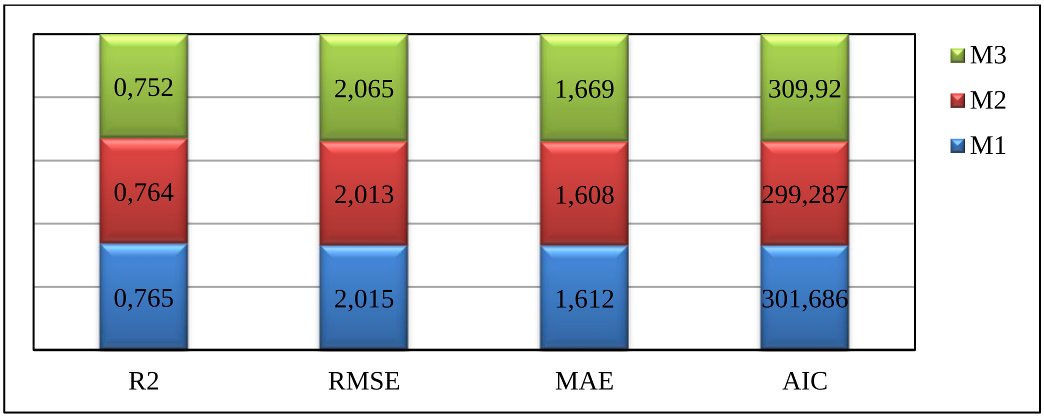

Figure 3. Statistical criteria for the models

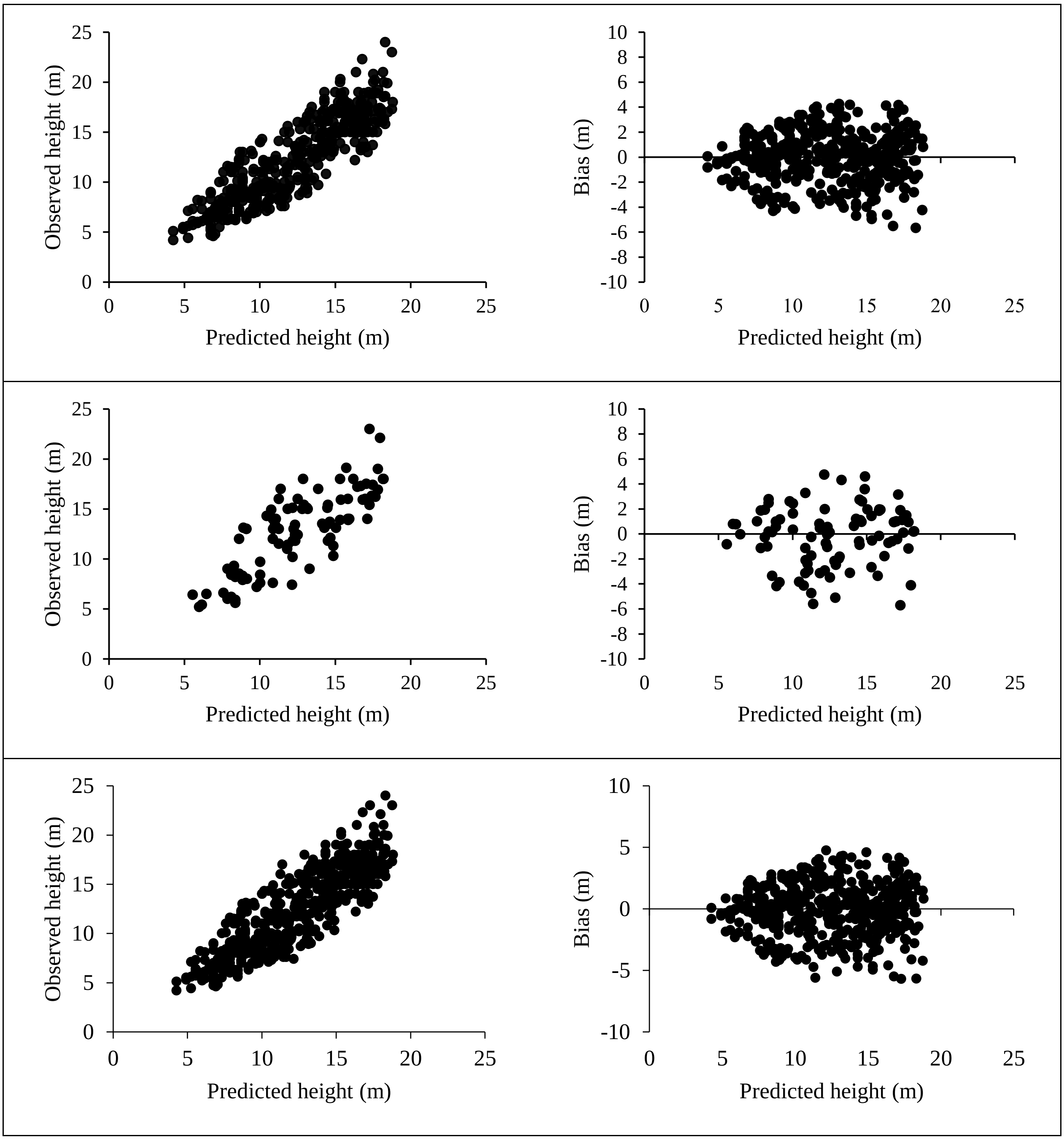

Figure 4. Prediction and bias plots of M2 for the model (top), control (middle) and total (bottom) data

In Türkiye, height–diameter relationships were studied by Sönmez (2009), Mısır (2010), Özçelik et al. (2013), Özçelik and Çapar (2014), Ercanlı (2015), Çatal and Carus (2018), Özçelik et al. (2018), Ercanlı and Eyüboğlu (2019), Ercanlı (2020a, 2020b), and Seki and Sakici (2022). Among these studies, the study conducted by Özçelik and Çapar (2014) was conducted in the Antalya province. Since the developed height-diameter model is a generalized height-diameter model, the predictions of the models were not compared. Height–diameter models are important because they serve as the basis for many other models. For this reason, studies on this topic are crucial. The height–diameter model developed for Brutian pine stands in the Kurşunlu Waterfall Nature Park stands out as a study that enables biomass estimation and carbon balance estimation in protected areas.

4. Conclusion

The statistical analyses were carried out for the development of height–diameter models for Brutian pine stands in the Kurşunlu Waterfall region in Aksu distirct of Antalya province within the scope of the study. Height–diameter measurements from 512 sample trees were used to develop the models. In the development of the models, three different height–diameter models were used, which are frequently preferred in the literature for modeling height–diameter relationships and give successful results. Four statistical criteria (R^2, *RMSE*, *MAE*, *AIC*) were used to determine the best model. Among the models evaluated, the two-parameter M2 model (Meyer, 1940) has been identified as the best model. After deciding on the best model, the validity of the model was tested using Student's *t*–test with an independent set of data, and as a result of this test, it was determined that the model provided unbiased predictions with 95% confidence. The height–diameter model developed for Brutian pine stands in the study area stands out as a study that allows biomass estimation and carbon stock estimation in protected areas. For this reason, it is an important study in terms of enabling carbon balance assessments for protected areas.

Acknowledgements

We would like to thank AKTÜR Mapping Office for providing the data for this study, and the administrative and technical staff of the 6th Regional Directorate of the Ministry of Agriculture and Forestry, especially Urban Planner Melisa BIYIKLI, for their support.

Conflict of interest

The authors declare that there is no conflict of interest.

Ethical Approval

For this type of study, formal consent is not required.

References

Arabatzis, A., & Burkhart, H. (1992). An Evaluation of Sampling Methods and Model Forms for Estimating Height-Diameter Relationships in Loblolly Pine Plantations. *Forest Science*, *38*(1), 192-198. https://doi.org/10.1093/forestscience/38.1.192

Çatal, Y., & Carus, S. (2018). A Height-Diameter Model for Brutian Pine (*Pinus brutia* Ten.) Plantations in Southwestern Turkey. *Applied Ecology and Environmental Research*, *16*(2), 1445-1459. http://dx.doi.org/10.15666/aeer/1602_14451459

Diamantopoulou, M., & Özçelik, R. (2012). Evaluation of Different Modeling Approaches for Total Tree-Height Estimation in Mediterranean Region of Turkey. *Forest Systems*, *21*(3), 383-397. https://doi.org/10.5424/fs/2012213-02338

Ercanlı, İ. (2015). Nonlinear Mixed Effect Models for Predicting Relationships between Total Height and Diameter of Oriental Beech Trees in Kestel, Turkey. *Revista Chapingo Serie Ciencias Forestales y del Ambiente*, *21*(2), 185-202. https://doi.org/10.5154/r.rchscfa.2015.02.006

Ercanlı, İ. (2020a). Innovative Deep Learning Artificial Intelligence Applications for Predicting Relationships between Individual Tree Height and Diameter at Breast Height. *Forest Ecosystems*, *7*, 12. https://doi.org/10.1186/s40663-020-00226-3

Ercanlı, İ. (2020b). Artificial İntelligence with Deep Learning Algorithms to Model Relationships between Total Tree Height and Diameter at Breast Height. *Forest Systems*, *29*(2), e013. https://doi.org/10.5424/fs/2020292-16393

Ercanlı, İ., & Eyüboğlu, D. (2019). Ağaçların Çap-Boy Modellemesine İlişkin Otokorelasyon Probleminin Giderilmesinde Karışık Etkili Doğrusal Olmayan Regresyon Modelleri ile Otoregresif Regresyon Modellerinin Karşılaştırılması. *Anadolu Orman Araştırmaları Dergisi*, *5*(1), 17-27. (in Turkish)

Huang, S., Price, D., & Titus, S. (2000). Development of ecoregion-based height–diameter models for white spruce in boreal forests. *Forest Ecology and Management*, *129*(1-3), 125-141. https://doi.org/10.1016/S0378-1127(99)00151-6

Larsen, D. R. & Hann, D.W. (1987). *Height–diameter equations*. Research paper 49, Forest Research Laboratory, Oregon State University, Corvallis.

Meng, S. X., Huang, S., Yang, Y., Trincado, G., & Vanderschaaf, C. (2009). Evaluation of Population-averaged and Subject-specific Approaches for Modeling the Dominant/Co-dominant Height of Lodgepole Pine Trees. *Canadian Journal of Forest Research*, *39*(6), 1148-1158. https://doi.org/10.1139/X09-039

Mısır, N. (2010). Generalized Height-Diameter Models for *Populus tremula* L. stands. *African Journal of Biotechnology*, *9*(28), 4348-4355.

Özçelik, R., Diamantopoulou, M. J., Crecente-Campo, F., & Eler, Ü. (2013). Estimating Crimean Juniper Tree Height Using Nonlinear Regression and Artificial Neural Network Models. *Forest Ecology and Management*, *306*, 52-60. https://doi.org/10.1016/j.foreco.2013.06.009

Özçelik, R., & Çapar, C. (2014). Antalya Yöresi Doğal Kızılçam Meşcereleri İçin Genelleştirilmiş Çap-Boy Modellerinin Geliştirilmesi. *SDÜ Orman Fakültesi Dergisi*, *15*(1), 44-52. (in Turkish)

Özçelik, R., & Kalkanlı, Ş., (2018). Kaş Yöresi Doğal Kızılçam (*Pinus brutia* Ten.) Meşçereleri İçin Ağaç Hacim Denklemlerinin Geliştirilmesi. *Turkish Journal of Forestry*, *19*(1), 9-19. (in Turkish)

Seki, M., & Sakici, O. E. (2022). Ecoregion-based Height-Diameter Models for Crimean Pine. *Journal of Forest Research*, *27*(1), 36-44. https://doi.org/10.1080/13416979.2021.1972511

Sharma, M., & Parton, J. (2007). Height-Diameter Equations for Boreal Tree Species in Ontario Using a Mixed-effects Modeling Approach. *Forest Ecology and Management*, *249*(3), 187-198. https://doi.org/10.1016/j.foreco.2007.05.006

Sönmez, T. (2009). Generalized Height-Diameter Models for *Picea orientalis* L. *Journal of Environmental Biology*, *30*(5), 767-772.

UDGP (2011). Kurşunlu Şelalesi Tabiat Parkı Uzun Devreli Gelişme Planı. Tarım ve Orman Bakanlığı 6. Bölge Müdürlüğü. (in Turkish)

Wang, C.H., Hann, D.W. (1988). *Height-diameter equations sixteen tree species in the central western Willamette valley of Oregon*. Research paper 51, Forest Research Laboratory, Oregon State University, Corvallis.

AUTHOR GUIDELINES

For submission, manuscript text should…

- be written in English.
- be written by MS-Word in A4 size.
- be margined 3-cm on top, 2-cm on bottom, and 1.5-cm on left and right sides.
- include page numbers (bottom, centered, in Arabic style).
- include continuous line numbers.
- be single-column and justified.
- be written in Times New Roman with 11-font size, except title, and 1.5-line spaced. Title should be written with 14-font size.
- include the following pages, separately:
  - Title page with authors' name(s), e-mail address(s), affiliation(s), abstract and keywords,
  - Main text with sections/subsections, figures, tables, references, and appendices (if any).

For publication, after the review process, manuscript text should…

- be double-column, except title, abstract and keywords. Title, abstract and keywords should be written in single-column.
- be justified, except title. Title should be centered.
- be arranged as: Title, Author(s) name(s) and affiliation(s), Abstract and Keywords, Main text (Introduction, Material and Methods, Results, Discussion, and Conclusion), Acknowledgements, Conflict of interest, and References.
- be written in single-spaced.
- be written in Times New Roman with 10,5-font size, except Title. Title should be written with 14-font size. Line spaces should be as: 24-pt above&below for title, 12-pt below for author name(s), and 6-pt below for other parts of the manuscript.

Title

The title of the manuscript should be informative and brief. It should not be too long or contain unnecessary words.

Abstract and Keywords

The abstract should be a single paragraph with a maximum of 300 words. The study rationale, objectives, methods, and findings should be briefly described in this part. Citations, tables or figures should not be included.

Keywords should be decided to optimize search engine discovery. They should be different from the words used in title.

Tables and Figures

Each figure and table should be cited in order of first appearance in the text (Figure 1, Table 1, etc.). Tables should be captioned at the top of the table in justified, while figures' captions should be given at the bottom of the figure with centered. Abbreviations and symbols used in the tables and figures should be defined in the text. Tables should be given without vertical lines and shading. Figures should be provided in TIFF or JPEG format.

Equations

Equations should be written by using a math editor (MathType) or LaTex. They should be identified with numbers in parentheses placed flush with the right margin. Each symbol within equations should represent only one entity.

Units of measure

The SI (International System of Units) units should be used. Decimals should be notated with a full stop.

Footnotes

Footnotes should be avoided in main text. If necessary, in text, they should be cited using superscript Arabic numbers and be given at the bottom of the page. In tables, footnotes should be cited using symbols.

Citations

The citations in the text should be given in APA Style. https://apastyle.apa.org/

Narrative citation:

Sakici (2020), Unal and Sakici (2020), Bayram et al. (2021)

Parenthetical citation:

(Sakici, 2020), (Unal & Sakici, 2020), (Bayram et al., 2020), (Sakici, 2020; Bayram et al., 2021)

References

The references should be written in APA Style. https://apastyle.apa.org/

Journal article:

Sakici, O. E., Saglam, F., & Seki, M. (2018). Single- and Double-entry Stem Volume Equations for Crimean Pine Stands in Kastamonu Regional Directorate of Forestry. *Turkish Journal of Forestry*, *19*(1), 20-29. https://doi.org/10.18182/tjf.394876

Book:

Husch, B., Beers, T. W., & Kershaw, J. A. (2003). *Forest mensuration* (4th ed.). Wiley.

Chapter in an edited book:

Köhl, M. (2004). Forest inventory and monitoring. In J. Burley, J. Evans, & J. A. Youngquist (Eds.), *Encyclopedia of Forest Science* (pp. 403-409). Elsevier.

Conference Presentation or Proceeding:

Sakıcı, O. E., & Gulsunar, M. (2012). *Diameter distributions of Bornmullerian fir in mixed stands* [Oral presentation]. 14th International Fir Symposium, Kastamonu, Turkey.

Dissertation or Thesis:

Saglam, F. (2016). *Constructing of above-ground biomass tables and developing of compatible biomass-volume equations for black pine (Pinus nigra J.F. Arnold) stands in Taskopru Forest Enterprise* [Master's thesis, Kastamonu University].

Web:

General Directorate of Forestry (2013). Forest atlas. https://www.ogm.gov.tr

Acknowledgements

The funding organization(s) and/or person(s) who indirectly contribute to the study should be thanked. Corresponding author is responsible for ensuring that the organization(s) and/or persons given in the Acknowledgements agree to be named.

SCOPE

Journal of Biometry Studies (JofBS) publishes articles on statistical and mathematical methods in life and natural sciences (health, biotechnology, agriculture, forestry, environment, wildlife, aquaculture, fisheries etc.). Research and review articles, technical reports, short communications, case studies and letters to the editors about aforementioned disciplines are welcome.

PUBLICATION ETHICS

Journal of Biometry Studies (JofBS) follows certain ethical standards for publication, existing to ensure high-quality scientific publications, public trust in scientific findings, and due credit for original ideas. JofBS is connected to the Committee on Publication Ethics (COPE), abides by its Code of Conduct, and aims to adhere to its Best Practice Guidelines. Committee on Publication Ethics (COPE). (2011, March 7). Code of Conduct and Best-Practice Guidelines for Journal Editors. Retrieved from https://publicationethics.org/about/guide/journal-editors

Author who submits any paper to JofBS certify that their work is original and is not published or under publication consideration elsewhere. Also, authors confirm that submitted papers have not been copied or plagiarized, in whole or in part, from other papers or studies. Authors certify that he/she does not have potential conflicts of interest or partial benefits associated with his or her papers.

JofBS will check for plagiarism in all submitted articles prior to publication. If plagiarism is detected at any stage of the publication process, the author will be instructed to rewrite the manuscript. Every submission will be scanned by any plagiarism software to prevent plagiarism. If any manuscript has over 20% similarity, the article will be rejected and the author will be notified. We strongly recommend the authors to check the similarity ratio of manuscript before submitting. Plagiarism can be checked by using any free online software.

JofBS is committed to objective and fair blind peer reviews of submitted papers and the prevention of any actual or potential conflicts of interest between writers and reviewers.

Responsibilities of the Editorial Board

All editors of JofBS are independent in their evaluations and decisions in the journal. No external and/or internal factor can affect their decisions. If the editors are exposed to any kind of positive and/or negative constraints, they keep the right to take legal action against those involved in the constraint. On the other hand, editors are responsible for their decisions in the journal. The editor-in-chief is the only person responsible for journal content and on-time publishing.

The members of Editorial Board are forbidden to share submitted materials with third parties other than section editors, language editors, copy editors, design editors and ombudsman when needed, and to use the submitted materials themselves. If there is a conflict of interest among an editor and an author or institution of the author in terms of cooperation or competition, then another member of the Editorial board is assigned to manage the evaluation process.

Editors provide peer review of submitted manuscripts by assigning at least two reviewers expert in the field. Editor-in-chief is responsible in decision of publishing a manuscript considering the importance of the manuscripts for researchers and readers, reviewer reports, plagiarism and copyright infringement as legal issues. Editor-in-chief can discuss with other editors and reviewers for his/her decision.

Responsibilities of Reviewers

Peer-reviewing of a submitted manuscript is the control of its scientific content, scientific layout and suitability according to the principles of the journal, and delivery of the reviewer's opinion for unsuitable manuscript content to ensure suitability. The reviewing process, not only enables reviewers to forward their evaluations about the manuscripts to the editors but also give them the opportunity to improve the contents of the manuscripts.

If a reviewer assigned for evaluation of a manuscript is of an expert in a field of science other than the manuscript content, is far to the subject of the manuscript, is short of time for evaluation or possess a conflict of interest, then he/she should inform the assigning editor and ask his/her withdrawal. If the content of the manuscript fits the expertise field of the reviewer, then he/she should complete the evaluation and send the report to the editor before the deadline.

Reviewers assigned for evaluation of manuscripts approve in advance that the manuscripts are secret documents and do not share any information about these documents with third parties except the editors involved in the evaluation. Reviewers continue to not to share information even after the manuscripts are accepted or rejected for publication.

If it is suspected of using an idea in the manuscript that is sent for evaluation to the reviewer without permission, the flowchart of COPE "What to do if you suspect a reviewer has appropriated an author's ideas or data?" is followed.

Reviewers should construct their criticisms on a scientific background and include scientific evidences in their statements. All comments raised by the reviewers to improve the manuscripts should be clear and direct and written in a manner far away from disturbing author's feelings. Insulting and derogatory statements should be avoided.

Reviewers should determine quotations in the manuscripts used without citing a reference. Statements, observations, conclusions or evidences in published articles should be quoted with the citation of the related reference. Reviewers should also be sure about the reality of presence of quotations in the cited reference(s).

If a reviewer is in a situation by being involved in one or more interests with the author(s), he/she should inform the editor the assigning editor and ask his/her withdrawal.

Responsibilities of Authors

Authors of original research articles should present the results and discuss with them in a proper way. Since the methodological contents of the articles should be reproducible, the authors should be clear in their statements and should not purposely report wrong or missing data. Authors of review type articles are not recommended to write such articles if they are not an expert in the field of their review topics or when they do not have enough background information or related former studies.

Authors may be asked to present their raw-data when needed (ethical cases etc.). Therefore, raw-data of the manuscripts should be kept in safety to present if needed. The storage period of raw-data following publications should be at least 10 years.

The authors of submitted manuscripts should be sure that their manuscripts are original or include cited references for quotations.

It is not an approved way to produce more than one publication reporting on the same research. The authors should pay attention to such cases and they should not submit the same manuscript to different journals simultaneously.

Only the following persons should be included in the manuscripts as responsible authors:

- Researchers providing major contribution to concept, design, performing, data collection and/or analysis in a study,
- Researchers involved in preparation or critical revision of manuscripts,
- Researchers approved the latest version of the manuscripts and accepted its submission.

Contributors other than the above list (technical assistance, helpers in writing and editing, general contributions, etc.) should not be involved in the authors list but can be listed in acknowledgements section. The corresponding authors of manuscripts should provide the separate listing of contributors as authors and those to be involved in acknowledgements section.

Authors should clearly declare any kind of conflict of interests in their manuscripts. Absence of conflict of interests about the topic of the manuscripts should also be declared. The most common types of conflict of interests are financial supports, education or other types of funds, personal or institutional relations and affiliations. All sources of financial supports (with their grant or other reference numbers) of the studies should be declared.

Authors should not use personally obtained information (conversations, correspondences or discussions with bystanders) unless they have the permission of their sources. Information about private documents or refereeing of grant applications should not be used without the permission of the authorities providing the related service.

Ethical guidelines for the use of animals in research

JofBS endorses the ARRIVE guidelines (www.nc3rs.org.uk/ARRIVE) for reporting experiments using live animals. Authors and reviewers can use the ARRIVE guidelines as a checklist, which can be found at www.nc3rs.org.uk/ARRIVEchecklist

Manuscripts containing original research on animal subjects must have been approved by an ethical review committee. The project identification code, date of approval and name of the ethics committee or institutional review board must be cited in the Methods section.

For research involving animals, any potentially derived benefits must be significant in relation to harm suffered by participating animals. Authors should particularly ensure that their research complies with the commonly accepted "3Rs":

- Replacement of animals by alternatives wherever possible,
- Reduction in number of animals used, and
- Refinement of experimental conditions and procedures to minimize the harm to animals.

OPEN ACCESS POLICY

JofBS is an open access journal publishing high quality papers that are original research and review articles, technical reports, short communications, case studies and letters to the editors. All published papers are freely available and openly accessible. The journal does not charge any article submission, processing or publication fees.

JofBS follows the guidelines presented by the Budapest Open Access Initiative (BOAI) regarding Open Access. It means that articles published in JofBS have free availability on the public internet, permitting any users to read, download, copy, distribute, print, search, or link to the full texts of these articles, crawl them for indexing, pass them as data to software, or use them for any other lawful purpose, without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself.

Please visit the given links below for more information about the Budapest Open Access Initiative.

https://www.budapestopenaccessinitiative.org/read

https://www.budapestopenaccessinitiative.org/boai-10-recommendations

https://www.budapestopenaccessinitiative.org/boai15-1

REVIEW PROCESS

Authors are obliged to be involved in the peer review process and should cooperate by responding raw data, evidence for ethical approvals, patient approvals and copyright release form requests of editors and their explanations. Authors should respond either in a positive or a negative way to revision suggestions generated by the peer review process. They should be sure to include their counter views in their negative responses.

Submitting authors must confirm the following:

- Manuscripts must be the original work of the submitting author.
- Submitted manuscripts must be unpublished.
- There should be no conflict of interest. If it exists, it must be clearly stated.
- Authors should cite all data sources used in the preparation of the manuscript.

Please note: It is unethical to submit a manuscript to more than one journal concurrently.

Reviewers must confirm the following:

- Manuscripts are reviewed fairly based on the intellectual content of the paper regardless of gender, race, ethnicity, religion, citizenship or political view of the author(s).
- Any observed conflict of interest during the review process must be sent to the editor.
- Information pertaining to the manuscript is kept confidential.
- Information that may be a cause for rejection of publication must be sent to the editor.

Editors must confirm the following:

- Manuscripts are reviewed fairly based on the intellectual content of the paper regardless of gender, race, ethnicity, religion, citizenship or political view of the author(s).
- Information pertaining to manuscripts is kept confidential.
- Any observed conflict of interest pertaining manuscripts must be disclosed.

DISCLAIMER

Editor or members of the editorial board are not responsible for the author's opinions and manuscript contents. Authors are responsible for the ethical originality of and possible errors in their manuscripts. They are also responsible for all errors based on page editing before their proofreading. On the other hand, errors taking place after proofreading are in responsibility of the journal directors.

LICENSE

Authors are able to enter into separate, additional contractual arrangements for the non-exclusive distribution of the journal's published version of the work (e.g., post it to an institutional repository or publish it in a book), with an acknowledgement of its initial publication in this journal.

Authors are permitted and encouraged to post their work online (e.g., in institutional repositories or on their website) prior to and during the submission process, as it can lead to productive exchanges, as well as earlier and greater citation of published work (see The Effect of Open Access).

www.biometrystudies.com
www.prensip.gen.tr

2024