Data analysis of treated wastewater impacts on coastal water parameters

Dimitrios Kazolis [1*], *Ioanna* Petrakopoulo[1], *and Theodosia* Kazoli[2]

[1]Democritus University of Thrace, Department of Physics, 65404 Kavala, Greece
[2]University of Macedonia, Department of Applied Informatics, 54636 Thessaloniki, Greece

Abstract. The discharge of treated or partially treated wastewater into water bodies is a common practice in many modern cities. However, this has often led to water quality degradation, ecological damage, and, in many cases, potential public health effects. This paper examines whether treated wastewater, discharged into a river, can affect coastal water quality at the river's mouth, located more than 80 kilometers away. For this purpose, measurements were taken over a period of five years, covering both the chemical characteristics of the wastewater and data on the water quality of the coastal ar-ea near the river mouth. After being properly processed, these data were combined based on their common date. The relationship between the variables was then analyzed to reveal potential correlations. Initially, an association was observed between the data from the two datasets, mainly through weak to moderate correlations between specific parameters. The nature of this ef-fect suggests limited direct linear relationships between many environmental indicators from the two sources. For this reason, further investigation was performed by applying Granger Causality test, multivariate Granger Causality and time-lagged correlations, especially for parameters with weak current linear correlations, to uncover more complex dependencies. Finally, the results were visualized and analyzed to draw useful conclusions.

1 Introduction

Industrial, urban, and other types (i.e. agricultural) of wastewater, are discharged from their facilities as effluent that is released into the environment, affecting the quality of the aquatic ecosystem. The concentration of the pollutants released into the environment must comply with regulations, and legislation limits set by each prefecture in Greece. Effluents coming from every industry, whether treated or not, must not exceed the specific limits set for them. Water discharged from wastewater plants can increase phytoplankton growth and alter coastal marine ecosystems, so water quality indicators have been studied to evaluate the environmental impact and correlate changes with other parameters [1, 2].

Wastewater treatment processes have advanced over the years. Nevertheless, the various parameters of the effluents need to be studied in order to assess the consequences they impose on the environment. The raw wastewater of the treatment plants is the influent, whereas the

* Corresponding author: dkazolis@physics.duth.gr

treated wastewater is the effluent. The evaluation of the central industrial water treatment performance of Larissa, has been performed in 2011 by Tzoupanos and his team [3], where it was found that several effluent parameters exceeded the prefectural legislation limits. In another study back then, from Bakopoulou et.al [4], the effluents from four wastewater plants in Thessaly region were found to be acceptable for reuse in irrigation.

The discharged water taken into account for this study consists of the effluents from two major wastewater plants: the CTU of the industrial area of Larissa, located in central Greece, in the region of Thessaly, and the urban wastewater treatment plant of the city. The effluent treated water from these plants is discharged into the river of the region (Pineios River), which flows for several kilometers to eventually run into the sea. The industrial facilities include pharmaceutical, textile dye, food, and material processing factories. Regarding the industrial CTU, there have been some upgrades to the facility recently. However, it must be mentioned that there are several factories include their own wastewater treatment plants, and their effluent activity is not included in this da-ta-processing study.

The parameters examined here included the effluent values of BOD, COD, pH, DO, TP, TS, nppv, SS, NO_3-N, NH_4-N, chlorofyll, and EC. It must be noted that each parameter referring to the industrial effluent value and to the urban effluent value of the corresponding wastewater treatment plants was examined separately, and not as a sum of the two wastewater treatment units. These parameters were compared and studied in correlation with the seawater parameters of the coastal region near the river delta.

2 Materials and methods

As mentioned above, the data were collected from separate sources, namely the industrial area of Larissa, the wastewater treatment plant, and the Copernicus website [5]. In the first phase, these were subjected to all the necessary procedures [6, 7], so that they could then be processed more accurately, e.g., selection of specific dates, removal of empty and extreme values, etc.

2.1 Correlation analysis

After this preparation, a correlation analysis was performed. The correlation matrix between the 'sea measurements' columns and the 'city effluents' columns was calculated to understand their relationships. The above procedures were performed using the Python programming language. Thus, to visualize the correlation matrix, the necessary plotting libraries, such as `matplotlib.pyplot`, `seaborn`, and `seaborn.heatmap`, were imported to generate the heatmap and highlight the strength and direction of the relationships between effluent parameters and sea. This matrix is presented in the following Figure 1.

By examining the correlation matrix, a number of observations can be made:

- For ph_1 (sea pH):

It indicates a moderate positive correlation with COD-mg/L (0.18) and TSS-mg/L (0.17), suggesting that an increase in these pollutants may be associated with a slight increase in sea pH.

Moreover, there is a notable positive correlation with NH_4-N (0.19) and TN-mg/L (0.18), indicating that increased nitrogen in effluents may affect seawater pH.

There is also a moderate positive correlation with DO-mg/L (0.25).

Finally, it has a moderate negative correlation with pH_eff (-0.31), meaning that as the pH of effluents increases, the sea pH tends to decrease.

- For nppv_1 (primary productivity):

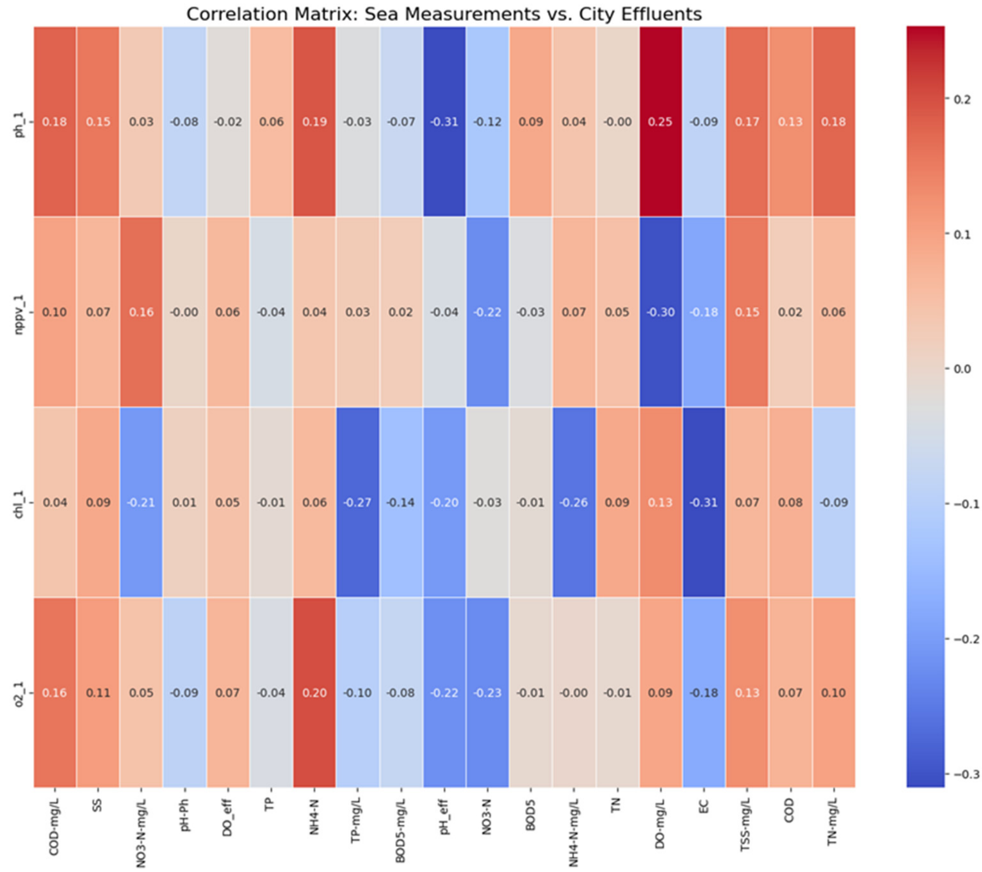

Fig. 1. The correlation matrix.

This indicates a moderate negative correlation (-0.30) with Dissolved Oxygen (DO-mg/L, suggesting that increased dissolved oxygen in effluents may be associated with decreased primary productivity in the sea.

Furthermore, it has a small positive correlation (0.10) with Chemical demanded oxygen (COD-mg/L) and (0.15) with Total Suspended Solids (TSS-mg/L), which might indicate that increased organic matter and suspended solids in effluents could influence primary productivity.

- For chl_1 (chlorophyll):

Here, there is a negative correlation (-0.21) with Nitrate-Nitrogen, (NO_3-N-mg/L), and (-0.26) with ammonium Nitrogen (NH_4-N-mg/L), suggesting that increased nitrogen in effluents may be associated with reduced chlorophyll levels.

It also has a negative correlation (-0.27) with Total Phosphorus, (TP-mg/L) and (-0.31) with Electrical Conductivity (EC).

- The o2_1 (oxygen):

It has a moderate positive correlation with NH_4-N (0.20), suggesting that increased ammonium nitrogen in effluents may be associated with increased oxygen levels in the sea. It also shows positive correlations with ph_1 (0.85), chl_1 (0.69), and (0.51) in Net Primary Production Velocity (nppv_1), which are expected as these are correlations among the sea measurements themselves. Regarding the impacts from urban effluents, there is a small positive correlation with COD-mg/L (0.16) and (0.11) Suspended Solids, (SS), which may indicate that organic matter and suspended solids in effluents have a small positive relationship with sea oxygen, although this would need further investigation.

The above analysis indicates some evidence that urban effluent pollutants may affect sea water quality. For example, sea pH appears to be influenced by effluent pH, while chlorophyll and primary productivity show negative correlations with some effluent pollution indicators.

However, it is important to mention that "correlation does not imply causation" and that the fact that if two variables move together (are correlated), this does not necessarily mean that one causes the other [8]. For this reason, although this analysis highlights potential relationships, further research, using time-series analysis, is needed to confirm causal links and the mechanisms of impact.

2.2 Time series analysis

2.2.1 Granger causality test

For the above reason, time-series analysis, specifically the Granger causality test, can provide more specific indications of causality in this endeavor. To prepare the time series data for analysis, the 'date' column was converted to a datetime object, set it as the Data-Frame's sorted index, and it was ensured that all time series data were aligned. Also, the missing values of NaNs were handled using appropriate interpolation methods. Finally, stationarity tests were performed on all time series of sea and wastewater measurements, and differencing was applied to make the series stationary, which is necessary for the Granger Causality test [9].

Following the above, the Granger Causality test was applied to determine whether the wastewater parameters "cause" the marine quality measurements. The appropriate number of lags for the test was selected as five.

For reasons of space, only two of the recordings are presented in Figure 2 and Figure 3.

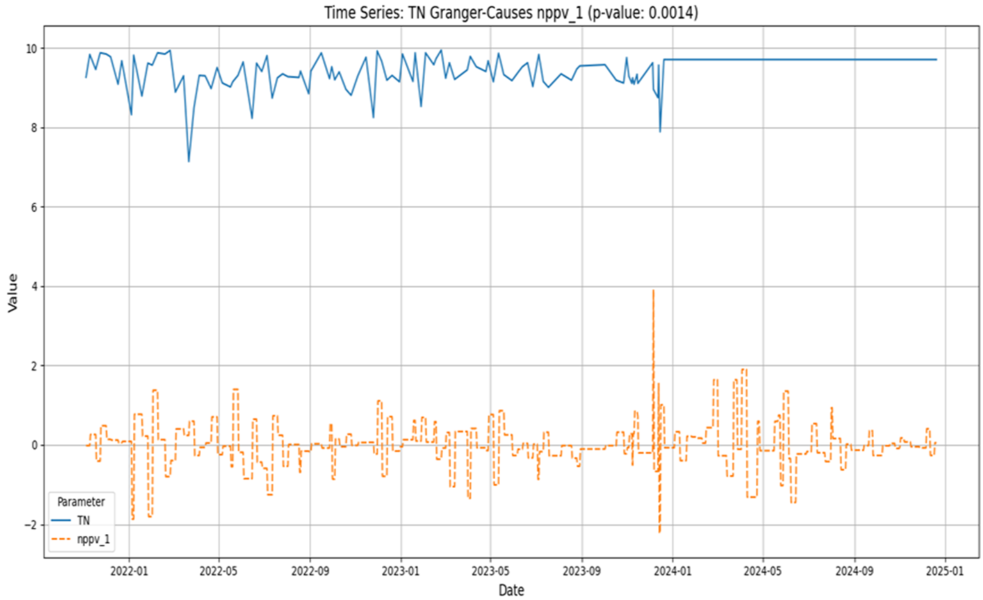

Fig. 2. Time Series. TN "Granger-causes" nppv_1.

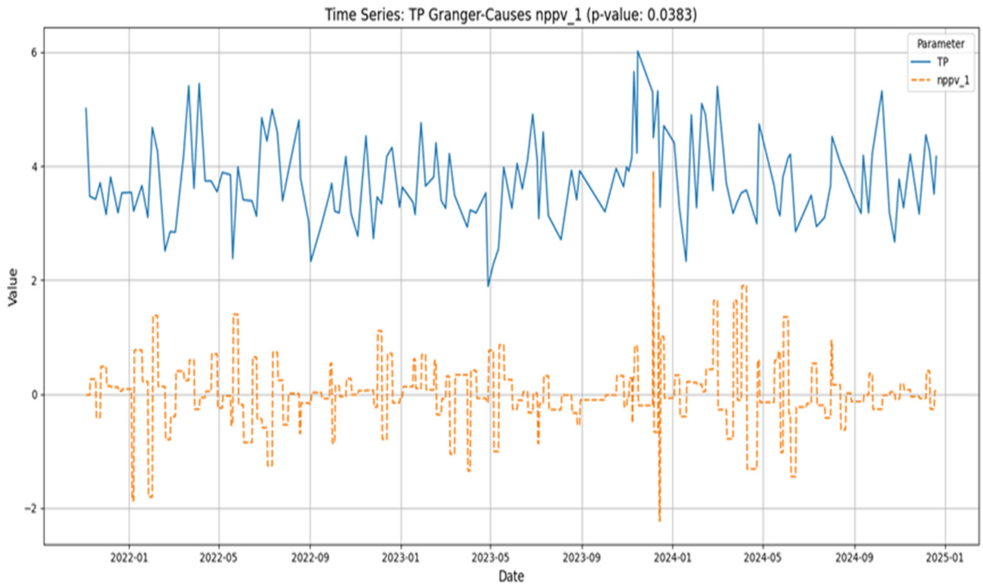

Fig. 3. Time Series. TP "Granger-causes" nppv_1.

However, it should be stated that results were also obtained for the "Granger-cause" of COD and nppv_1 with p-value 0.0249, DO and o2_1 with p-value 0.0172, NO_3-N and o2_1 with p-value 0.0309, TN and o2_1 with p-value 0.0205, BOD5 and ph_1 with p-value 0.00, COD and ph_1 with p-value 0.00, SS and ph_1 with p-value 0.0031, NH_4-N and ph_1 with p-value 0.0434, DO_eff and ph_1 with p-value 0.0268, BOD5-mg/L and ph_1 with p-value 0.0322, DO-mg/L and ph_1 with p-value 0.0037, and EC and nppv_1 with p-value 0.0322.

In summary, Granger Causality tests with max_lag= 5 revealed significant relation-ships ($p < 0.05$), and the following conclusions can be drawn:

• For seawater pH (ph_1): Biochemical demanded oxygen (BOD) with p-value = 0.000002 and BOD5-mg/L` with p-value = 0.0322, Chemical demanded oxygen (COD) with p-value = 0.000001, Suspended Solids (SS) p-value = 0.0031, Ammonium-Nitrogen (NH_4-N) p-value = 0.0434, Dissolved oxygen (DO_eff) with p-value = 0.0268, and DO-mg/L with p-value = 0.0037, appear to Granger-cause the sea water ph_1.

• For Net Primary Production Velocity (nppv_1`): Total Nitrogen (TN) p-value = 0.0014, Total Phosphorus (TP) p-value = 0.0383, Electrical Conductivity (EC) p-value = 0.0322, and COD with p-value = 0.0249 from city effluents appear to Granger-cause the variable nppv_1.

• For Dissolved Oxygen in sea water (o2_1): Dissolved oxygen (DO) p-value = 0.0172, Nitrate-Nitrogen (NO_3-N) p-value = 0.0309, and Total Nitrogen (TN) p-value = 0.0205 appear to Granger-cause the oxygen (o2_1) in sea.

• For Chlorophyll-a (chl_1): No statistically significant Granger-causal relationships were found with any city effluent parameters at the 5% significance level.

The summary the above high significance values, are presented in the following Table1.

Nevertheless, it should be noted at this point that the Granger Causality test indicates predictive power but does not prove direct physical or chemical causation. It is a statistical relationship, and the underlying environmental and hydrological mechanisms need further investigation. Additionally, the choice of `max_lag`, set to 5 in this analysis, can in-fluence the results, and an optimal lag length might vary for different pairs.

For the above reason, it was deemed appropriate that the results of the Granger Causality test be compared with the initial correlation matrix.

Table 1. Granger causality tests.

Variable	p-value	Variable	p-value
BOD	0.000002	TN	0.0014
BOD5-mg/L	0.0322	TP	0.0383
COD	0.000001	EC	0.0322
SS	0.0031	COD(nppv_1`)	0.0249
NH_4-N	0.0434	DO	0.0172
DO_eff	0.02680	NO_3-N	0.0309
DO-mg/L	0.0037	TN (o2_1)	0.0205

The correlation matrix, in general, reveals the linear relationship between each pair of variables at a given point in time, without considering the temporal order of events. In this study, various positive and negative correlations were identified. For example, ph_1 had a positive correlation with COD-mg/L and NH_4-N, and a negative one with pH_eff. On the other hand, correlation does not imply causation, and as a result, it cannot define if the increase in one factor causes an increase in another, whether both are influenced by a third factor, or if it's a coincidental relationship.

On the other hand, the Granger Causality test examines whether the past values of one time series can be used to predict the future values of another time series. In the previous chapter, several significant Granger Causality relationships were found.

The main difference is that correlation is asynchronous, while Granger Causality examines the temporal direction of the effect. While a high correlation can be a sign of potential causality, Granger Causality provides a stronger argument for whether changes in an effluent factor can predict changes in seawater quality. So, it's interesting to note that some variables that had moderate correlations in the correlation matrix might not show Granger Causality, and vice versa.

For example, ph_1 had a negative correlation with pH_eff (-0.31) in the correlation matrix, suggesting they move in opposite directions. However, the Granger Causality test showed that pH_eff does not Granger-cause ph_1 (p-value = 0.0639, marginally non-significant at the 5% level). This indicates that, although there is a linear relationship, past values of effluent pH do not significantly improve the prediction of sea pH in this lag structure.

2.2.2 Multivariate granger causality

The Granger Causality test is a powerful tool for time series analysis, but it has some limitations, the most important of which is that it does not imply true causality. Further-more, it has sensitivity to lag selection, requires a stationary time series, assumes only linear relationships, has sensitivity to missing data and outliers, etc. For the above rea-sons, it was considered appropriate to apply the multivariate Granger Causality [10], which would allow us to examine the combined effect of many wastewater parameters on a sea measurement, considering the interactions between the wastewater parameters themselves and the marine measurements.

To achieve this goal, a Vector Autoregression (VAR) model was fitted to the combined stationary data on sea measurements and city effluents. Then, the VAR model was used to perform a multivariate Granger Causality test.

As a result, the multivariate Granger Causality test performed, using an optimal lag order of 2 (selected by AIC, up to a max-lag of 5), concludes that the null hypothesis, that the group of urban wastewater parameters is not a Granger cause for the group of marine measurement parameters, is rejected. This suggests that, overall, changes in urban wastewater parameters statistically precede and have predictive power in relation to changes in marine water quality measurements.

So, no contradiction between the two types of tests is observed. The pairwise tests showed numerous significant relationships, particularly for seawater pH, net primary production velocity, and dissolved oxygen in seawater, being Granger-caused by various effluent parameters (Figure 4). The multivariate test indicates that the entire set of city effluent parameters, when considered together, significantly influences the overall group of sea measurements. To visually compare the results a heatmap was created.

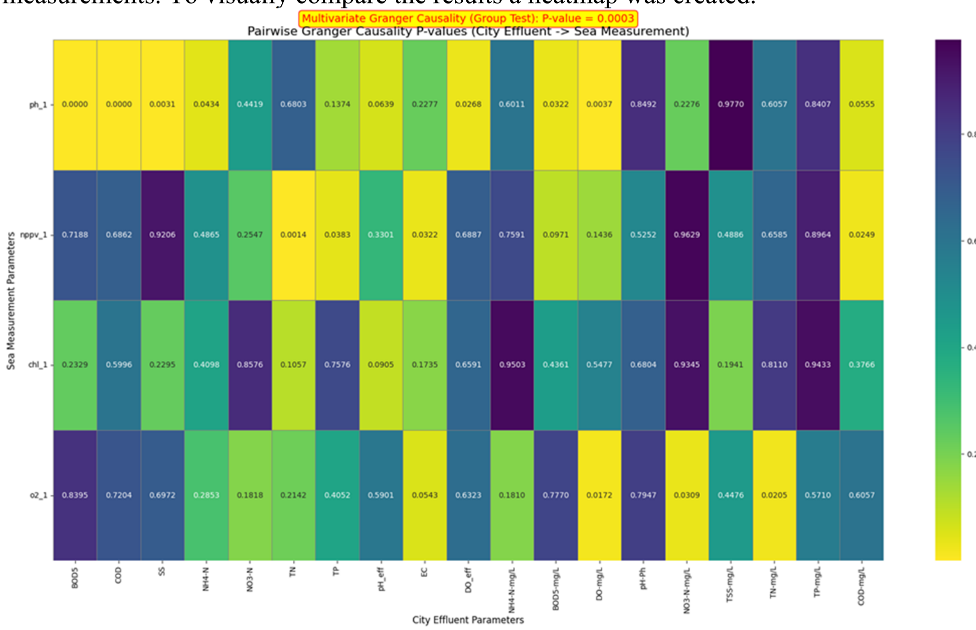

Fig. 4. Comparison of results: pairwise vs multivariate granger causality.

This heatmap, depicts the p-values for each 'cause-effect' pair from pairwise Granger Causality.

In this image, a note was added with the p-value of multivariate Granger Causality for direct visual reference. From this graph, it is immediately apparent which specific relationships are statistically highly significant ($p < 0.05$), as these will have very low p-values. The darker the colour of a cell, the smaller the p-value and the stronger the indication that the wastewater parameter Granger causes the marine measurement.

Moreover, the p-value shown at the top of the graph (P-value = 0.0003) is extremely low and suggests that, as a group, the urban wastewater parameters are Granger-causal to the marine measurements. This is a conclusion at the aggregate level and cannot be directly observed from individual cells of the heatmap.

2.3 Time-Lagged Correlation

As mentioned in the introduction, the waste is dumped into the Pineios River and, after flowing several kilometers, it is emptied into the sea. Thus, it was considered appropriate to examine the possibility of a time lag. So, the time-lagged correlation analysis was performed between each sea measurement parameter and each city effluent parameter, considering lags up to 14 days in both positive and negative directions. The goal was to identify significant leading or lagging relationships that might suggest potential causal pathways or influences. As an example, Figure 5 is presented.

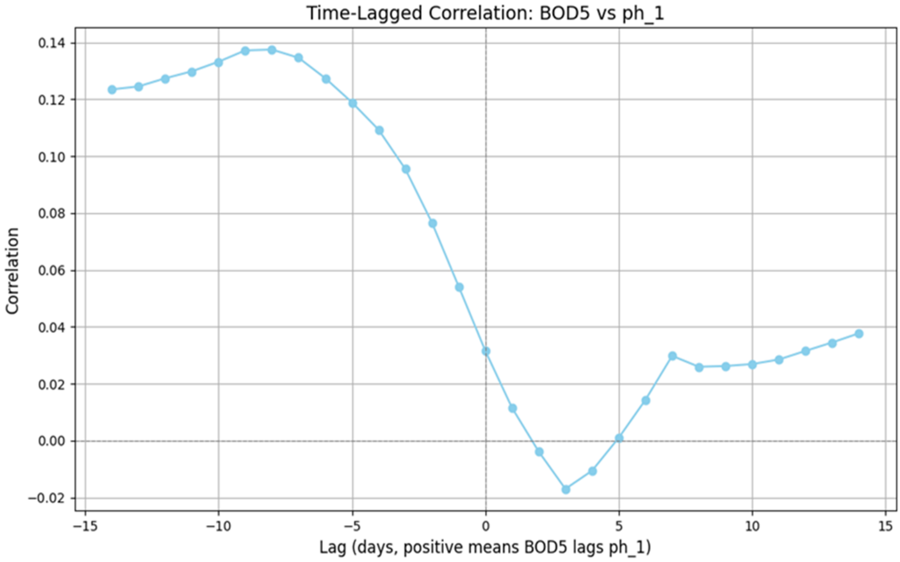

Fig. 5. Time -lagged correlation BOD5 vs ph_1.

The most significant relationships that emerged were the following:

• For the sea water pH (ph_1): Many city effluent parameters showed their strongest correlations with ph_1 at negative lags, indicating that changes in city effluents lead to changes in sea water pH. For example, in the previous image, BOD5 showed a peak correlation of around 0.14 at a lag of -9 days, suggesting that higher BOD5 levels in effluents are followed by changes in sea water pH about 9 days later. Similarly, COD and NH4-N also exhibited peak correlations at negative lags, implying that these organic and nutrient loads from effluents influence seawater pH over a period of several days. Some parameters also showed synchronous or slight positive lags, but the negative lags leading influence were more common for stronger correlations.

• For Net Primary Production Velocity (nppv_1): Total Nitrogen (TN) showed a notable correlation with nppv_1 at negative lags, suggesting that nitrogen loading from effluents might present changes in primary production in the sea. Total Phosphorus and Electrical Conductivity also had peak correlations at negative lags, indicating a leading effect on nppv_1. These findings align with the idea that nutrient input can stimulate primary production after a certain delay.

• For Chlorophyll (chl_1): Chlorophyll-a generally showed weaker time-lagged correlations compared to ph_1 and nppv_1 with city effluent parameters. When correlations were notable, they varied across lags. Some effluent parameters, such as certain nitrogen or phosphorus compounds, showed weak positive or negative lagged correlations, but no consistently strong leading patterns were immediately evident across multiple effluents. This might suggest that the relationship between city effluents and chl_1 is more complex, potentially non-linear, or influenced by other factors not captured by this lagged correlation.

• For Dissolved Oxygen in Sea Water (o2_1): DO-mg/L and Dissolved Oxygen in effluents (DO_eff), often showed their strongest correlations with o2_1 around lag zero or slight negative lags, indicating a relatively quick or synchronous relationship. This is expected as effluent DO directly impacts the receiving water body's oxygen levels. Moreover, NO_3-N-mg/L and TN-mg/L, also exhibited significant correlations, often at negative lags, implying that nitrate and total nitrogen in effluents can influence sea water dissolved oxygen over time, possibly through processes like decomposition of organic matter or algal blooms and subsequent decomposition.

3 Conclusion

In this study, different approaches were used to explore the relationship between the treated wastewater and coastal water parameters. The results of each of these methods were presented in detail in the previous section.

However, the general conclusions that could be drawn are as follows:

Firstly, the correlation analysis provides some evidence suggesting that treated urban wastewater may affect seawater quality. As can be observed in the Correlation Matrix, this analysis provides an initial overview of potential relationships. To support these findings, the Granger Causality test was then applied. This test reinforced the analysis by providing evidence of its predictive power and confirmed that changes in specific wastewater parameters can explain changes in seawater quality. Nevertheless, it still does not prove physical or chemical causation, but it provides a stronger statistical argument for further investigation into the underlying mechanisms.

The following multivariate Granger Causality test, which focuses on the collective effect, confirms and strengthens the trend observed in the pairwise Granger Causality tests. While pairwise tests identified specific effluent parameters Granger-causing individual sea measurements, the multivariate test aggregates these individual causal relationships into a single, strong collective impact.

Finally, from Time-Lagged Correlation Analysis, a dominant pattern observed was that the strongest correlations often occurred at negative time lags (e.g., from -14 to -1 days). This suggests that changes in urban wastewater parameters generally precede changes in marine water parameters, indicating that, due to the river, the impact of urban wastewater on seawater quality is noticeable with a time lag.

To continue the analysis and deepen the understanding of the relationship between urban wastewater and marine water quality, additional methods could be applied. For example, by optimizing lag selection, impulse response analysis, investigation of non-linear relationships by exploring non-linear causality tests, and integration of external factors by adding and analyzing additional environmental data.

Lastly, it should be noted that the values of the seawater parameters used in this study were obtained from the “Copernicus” program; therefore, a corresponding level of uncertainty and error must be considered.

References

1. J. Su, Y. Qiu, Y. Lu, X. Yang, S. Li, Use of Multivariate Statistical Techniques to Study Spatial Variability and Sources Apportionment of Pollution in Rivers Flowing into the Laizhou Bay in Dongying District. Water, **13**, 772 (2021)

2. Y. Yang, Y. Gulbahar, Spatiotemporal modeling of water quality trends in a coastal wildlife refuge: a statistical approach to ecological risk and resource management. Model. Earth Syst. Environ., **12**, 45 (2026)
3. N. D. Tzoupanos, P. Karamanolis, P. Zannis, A. I. Zouboulis, Evaluation of the treatment efficiency of the central treatment unit (CTU) of the industrial area of Larisa (Greece). Desalination and Water Treatment, **39**, 248-255 (2012)
4. S. Bakopoulou, C. Emmanouil, A. Kungolos, Assessment of wastewater effluent quality in Thessaly region, Greece, for determining its irrigation reuse potential. Ecotoxicology and environmental safety **74**, 188-194 (2011)
5. Copernicus Marine Service. Available online: https://marine.copernicus.eu/access-data/ocean-visualisation-tools (accessed on 28 March 2026)
6. Vlahavas, I.; Kefalas, I.; Bassiliades, P.; Kokkoras, N.; Sakellariou, F. In Book Artificial Intelligence, 4th ed., University of Macedonia Press. (2020)
7. D. Kazolis, Investigation of the impact of photovoltaic systems on medium voltage grids. In Proceedings of the 6th International Conference on Communications, Information, Electronic and Energy Systems. Ruse, Bulgaria, 26-28 November (2025). DOI:10.1109/CIEES66347.2025.11300070
8. N. Barrowman, Correlation, causation, and confusion. (The New Atlantis, 2014)
9. A. Shojaie, E. Fox, Granger Causality: A Review and Recent Advances. Annual Review of Statistics and Its Application, **9**, 289-319 (2022)
10. K. Blinowska, R. Kuś, M. Kamiński Deep Neural Networks for Multivariate Granger-Causal Feature Extraction. APS Physical Review E, **70** (2004)