PLS-PM FOR ECOLOGICAL INTEGRITY MAPPING: CASE STUDY OF THE AVE RIVER BASIN, PORTUGAL

ANTÓNIO FERNANDES[1], ANA FERREIRA[1], LUÍS SANCHES FERNANDES[1],
RUI CORTES[1] & FERNANDO PACHECO[2]
[1]CITAB – Centre for the Research and Technology of Agro-Environment and Biological Sciences,
Universidade de Trás-os-Montes e Alto Douro, Porgugal
[2]CQVR – Vila Real Chemistry Research Centre, Universidade de Trás-os-Montes e Alto Douro, Portugal

ABSTRACT

The bonds between pollution sources, contaminant concentrations and ecological damage are quite complex to access, but the proper management of river basins requires the full understanding of those interactions. In the present study, four SEM-PLS models were used to set up environmental cause–effect relationships in a heavily polluted urban catchment: the Ave River Basin. Data were collected within this watershed relative to point source discharges, diffuse emissions, diffuse indicators (e.g. land use), contaminant concentrations and a Portuguese index of macroinvertebrates diversity (IPtIN). The same dataset was used in the four SEM-PLS models, but each run was based on a different combination of latent variables. In all cases, it was possible to recognize that nitrogen discharges from livestock production, industrial and urban treatment plants are the dominant threats to water quality. The model results were robust, characterized by high, adjusted R Squared values – from 0.744 to 0.931. The models were also used to draw water quality maps. Now, the purpose was to test their prediction capability. Since more than 80% of the water courses were correctly classified for IPtIN, the predictions were considered feasible.

Keywords: SEM-PLS, water quality, ArcGIS, Ave River basin.

1 INTRODUCTION

Growing population, intense economic development and following demographic expansion place not only heavy demands but also pollution sources in hydric resources. To respond to such threats is crucial to apply conservative management policies, guided by supportive studies with an environmental scope.

In water quality studies is crucial to use models, which can be process-based (mechanistic) or data-based (statistical) [1]. For the first case, it is necessary to have a deep scientific awareness, since these models are based in physical, chemical and biologic processes [2]. Streeter-Phelps equation [3], is a clear example of a mechanistic model, where dissolved oxygen and oxygen demands can be calculated along with time and distances. Many process-based models can be used for surface water quality studies, since they reveal to be powerful tools that can be used to predict transport of contaminants in water bodies, such as SWAT (Soil and Water Assessment Tool) [4], WASP (Water Quality Analysis Simulation Program) [5], MIKE 11 [6], HSPF (Hydrological Simulation Program-FORTRAN) [7], ELCOM-CAEDYM (Estuary and Lake Computer Model Computational Aquatic Ecosystem Dynamics Model [8]. Data-based models appeal to statistical methods and can be used for prediction or to establish cause-effect relationships. The advantage is that for experimental datasets the interactions between variables are revealed, what does not happen in process-based models, where all the interactions between pollution sources and contaminants are already defined.

SEM (Structural Equation Modelling), is an advanced statistical tool that has been widely used mostly to reveal complex cause-effect relationships. Generically there are two types of SEM [9], CB-SEM (Covariance-based SEM) and PLS-SEM (Partial Least Squares SEM)

also called as PLS-PM (PLS Path Modelling). In the first case, the estimation procedure is based on a maximum likelihood estimation, while PLS-SEM is based in ordinary least squares regression [10]. The benefits of each type of SEM differs from study, purpose and dataset, but the opinion of the most appropriate method is still divergent [9], [10]. SEM first steps were given in social studies, nowadays non-social sciences are adopting this modelling technique, including environmental sciences. Already in 1994 SEM was applied to study surface water quality [11], years later other studies used SEM, for purposes such as an evaluation of groundwater quality [12], comparison of pollution sources in surface waters [13] and even population awareness about water quality [14], [15]. In a recent study [16] SEM-PLS was used to compare the relationship between point source pressures, diffuse pressures and diffuse indicators with contaminants and ecological integrity, while [17] used simplified SEM-PLS models.

In this study, SEM-PLS was applied to an Ave River basin environmental dataset. This area (Fig. 1) was chosen due to the vast water pollution background. In the second half of the 20th Century, it was classified as one of the most polluted river basins in Europe, due to the high effluent discharge without treatment. At the end of the century, the Portuguese Government was committed with a Plan to remediate the Ave River. Since then water quality has improved, but some problems persist, which has caught the attention of several experts and researchers that conducted studies to evaluate the pollution status and changes in this river basin. In 1992 a study [18] evidenced that several industries were the cause of heavy metal pollution. In 2009 when comparing data from previous studies [19] it was noticed that heavy metal pollution was decreasing [20], on the other hand, other threats have arisen in this river basin. The eutrophication effects were accessed in a study [21], accusing not only nutrients from discharges but even from agricultural practices, though other authors [22], [23] defend that livestock production is another threat for the water quality of this river basin. Therefore to approach such immensity of variables SEM-PLS was the chosen statistical method.
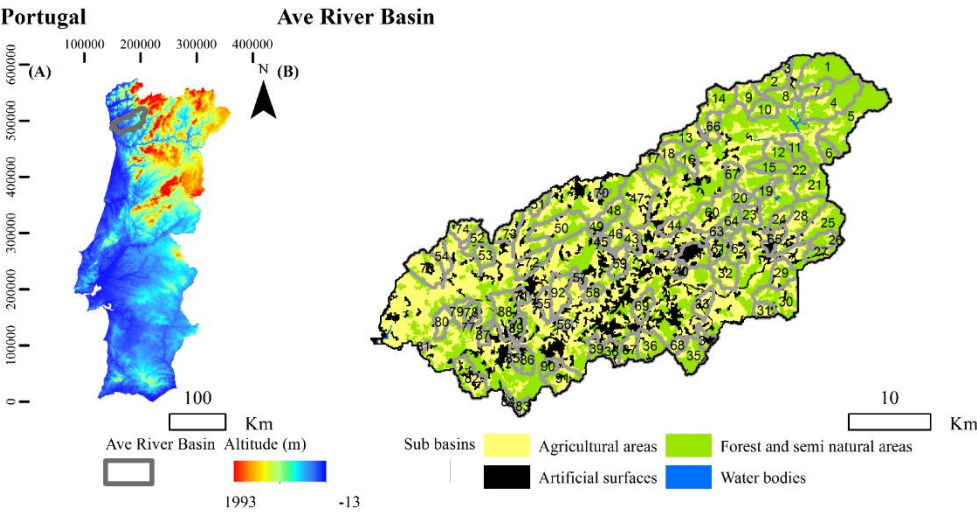

Figure 1:   Portugal Map with altitude, Ave River basin land uses.

2 METHODOLOGY

The objective of this study was to use pollution data in four structural equation models, transpose the results into water quality maps for Ave River basin. The applied six-step methodology is represented in Fig. 2.
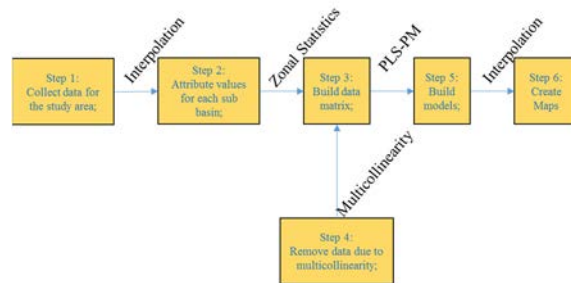

Figure 2: Schematic methodology.

As a first step, it was gathered all the necessary data for structural equation models. Data were grouped into three categories, pollution sources, contaminants concentration, and water quality.

For pressures, it was gathered a vast number of variables such as industrial and urban annual discharges in surface waters of phosphorous, nitrogen, oxygen demands (CBO_5 and COD) ($kg.year^{-2}$), percentage of area covered with different land uses and land use conflicts [24], population density, soil loss ($t.ha^{-1}.year^{-1}$), wildfire risk, nitrogen and phosphorous discharges from agriculture and forest (in the same variable) and livestock production ($kg.km^{-2}.year^{-1}$). For contamination, the gathered variables were the annual averages of concentrations of metals As, Cr, Cu, Fe, Pb, Zn ($\mu g/l$); nutrients NO_3 and PO_4 (mg/l); total suspended solids (mg/l) and oxygen demands COD and BOD_5 (mg/l). For Water Quality, was used as a single variable, North Invertebrate Portuguese Index ($IPtI_N$). This index was calculated through the count and diversity of benthic macroinvertebrates in river sites, variating the index score from 0 to 1 (dimensionless). A high score represents high macroinvertebrates biodiversity while a low score conveys that there is a lack of biodiversity. One advantage of the usage of this bioindicator is that macroinvertebrates are sensitive to practically all forms of pollution, so by calculating the $IPtI_N$ pollution effects are accessed, for a detailed explanation, please see [25]. All the data sources are demonstrated in Table 1.

For step 2, it was necessary to interpolate $IPtI_N$, surface water parameters, urban and industrial discharge, using Topo to Raster tool in ArcMap [26], to create a raster file for each variable. For other variables, it was not necessary to interpolate since the information was provided in the form of shapefiles and rasters covering the river basin. Since this study is hydrological and statistic, it was chosen to use sub-basins as statistical samples, delineated through ArcHydro [27]. In step 3 it was attributed a value of each variable for each sub-basin, using Zonal Statistic as Table tool, creating a data matrix for each river basin. In step 4 it was removed some variables due to multicollinearity, the used criteria were that the VIF (variance inflation factor) of each variable among a group (contamination, pressures) would not be higher than 5, which is a recommended limit value for SEM-PLS models [28]. For Water Quality this VIF analysis was not applied since this latent variable is composed of a single measured variable, $IPtI_N$. In Fig. 3 is demonstrated an example of an SEM-PLS formative model, composed of three latent variables (A, B and C) and nine measured variables (1 to 9).

Table 1: SEM-PLS data Source for models.

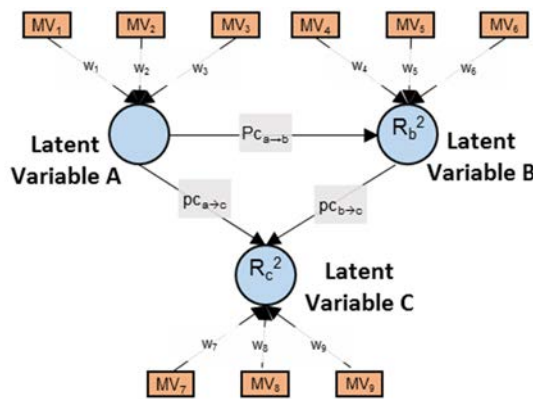| Variable	Used In Model	Units	Data Source
$IPtI_N$	x	Dimentionless	https://www.sciencedirect.com/science/article/pii/S014765131630207X
BOD_5	x	mg/l	http://snirh.pt/
COD	x	mg/l	
NO_3	x	µg/l	
PO_4	x	µg/l	
As		µg/l	
Cr	x	µg/l	
Cu		µg/l	
Fe	x	µg/l	
Pb		µg/l	
Zn		µg/l	
TSS		mg/l	
Wildfire Risk	x	dimensionless	http://www.icnf.pt
Population Density	x	inhabitant/km^2	http://censos.ine.pt
Agricultural Areas	x	%(km^2/km^2)	http://www.dgterritorio.pt/
Conflicts	x	%(km^2/km^2)	http://www.sciencedirect.com/science/article/pii/S0048969715313139
Agriculture and Forest N		kg/yr/ km^2	http://www.apambiente.pt
Agriculture and Forest P	x	kg/yr/km^2	
Livestock N	x	kg/yr/ km^2	
Livestock P		kg/yr/ km^2	
Industry BOD_5		kg/yr	
Industry COD	x	kg/yr	
Industry N	x	kg/yr	
Industry P		kg/yr	
Urban BOD_5		kg/yr	
Urban COD		kg/yr	
Urban N	x	kg/yr	
Urban P		kg/yr	

Figure 3: Formative structural equation model example.

As a fifth step, SEM-PLS models were created. In an SEM-PLS model, the user creates latent variables (LV), establishes the connections between them and chooses which measured variables (MV) compose each LV. The algorithm attributes weights (w) for each measured variable and path coefficients (pc) for each connection between LVs to maximize the R squared. For each LV is assigned to a measured and predicted score eqns (1) and (2)

$$\text{Measured Score: } LV_i = \sum_{i=1}^{n} (MV_i \times w_i), \tag{1}$$

$$\text{Predicted score: } LV_j = \sum_{i=1}^{n} (LV_{i \to j} \times pc_{i \to j}). \tag{2}$$

The measured score is based in the sum of the product of each MV with the attributed weight, while the predicted score is based in the sum product of the antecedent latent variables with the respective path coefficient. In the example, latent variable A is the only that does not have a predicted score because it does not have any antecedent latent variables. After determining the predicted score of $IPtI_N$ (for each sub-basin), it was created a raster file through the interpolation of the values in each sub-basin centroid. The created raster was used to attribute an average value of $IPtI_N$ for each watercourse, which is step 6.

3 RESULTS AND DISCUSSION

Four SEM-PLS models were applied for the same dataset. Model 1, Fig. 4, was depicted with only 3 latent variables, "pollution sources", "contamination" and "water quality", similar to one of the models presented in a previous study [17]. In Model 2 "water contamination" was divided into three types of contamination, by "oxygen demands", "nutrients" and "metals". Model 3 "pollution sources" were divided according to their type of source, "point source pressures", "diffuse emissions" and "diffuse indicators". Model 4, Fig. 5 it is the most complex model, where all the water contamination forms, and pollution sources are differentiated.

At first sight, all the models seem to be highly explicative, since the adjusted R squared values are high for "water quality", where the values for models 1 to 4 are 0.931, 0.816, 0.928 and 0.828, respectively. It is also stated that, for the same dataset, when more latent variables are used in a model, the algorithm loses the capacity to reach higher R-squared values for endogenous latent variables (with incoming arrows), because are shown the adjusted r-squared values in models.

Model 1

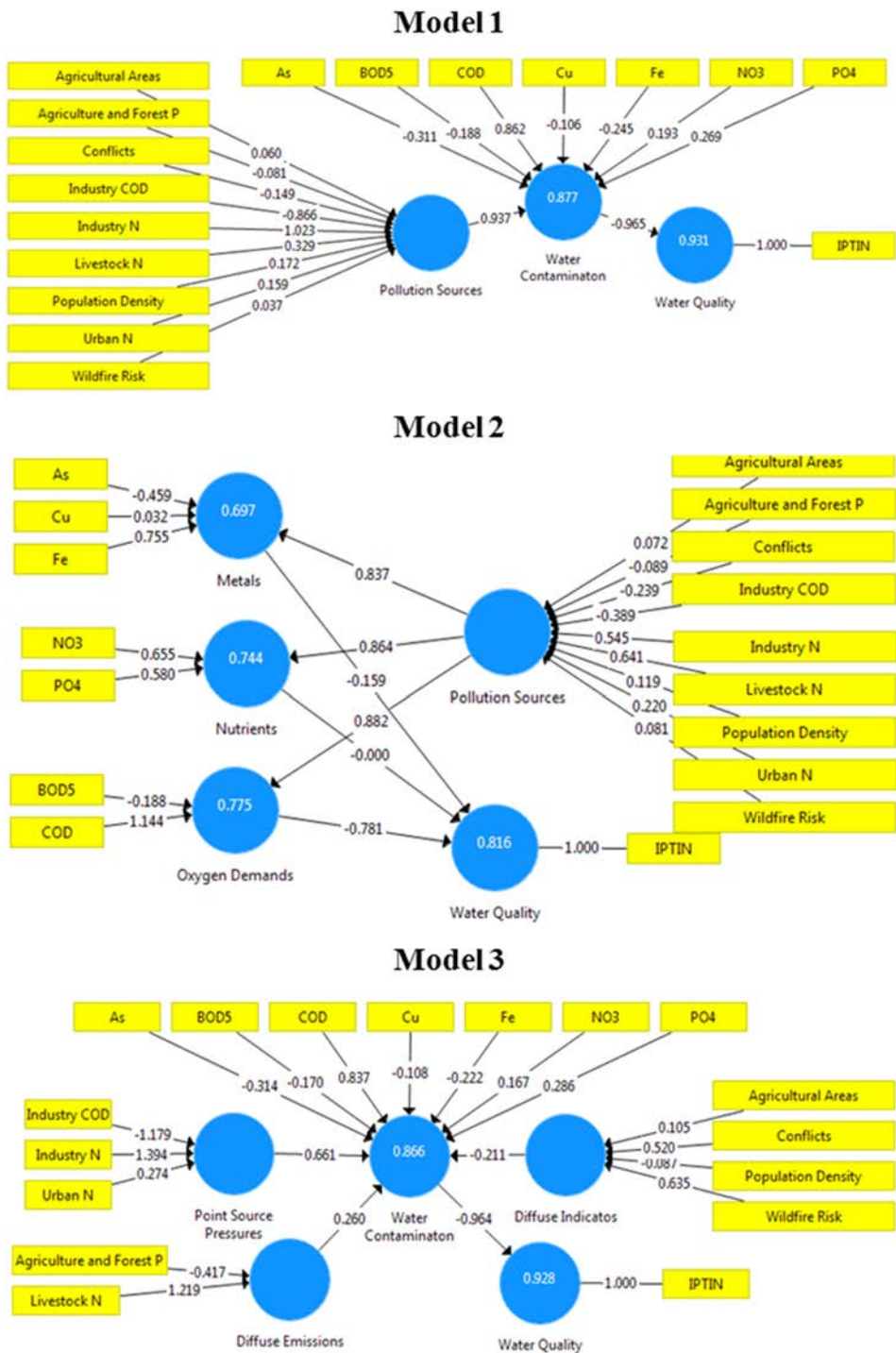## Model 2

Model 3

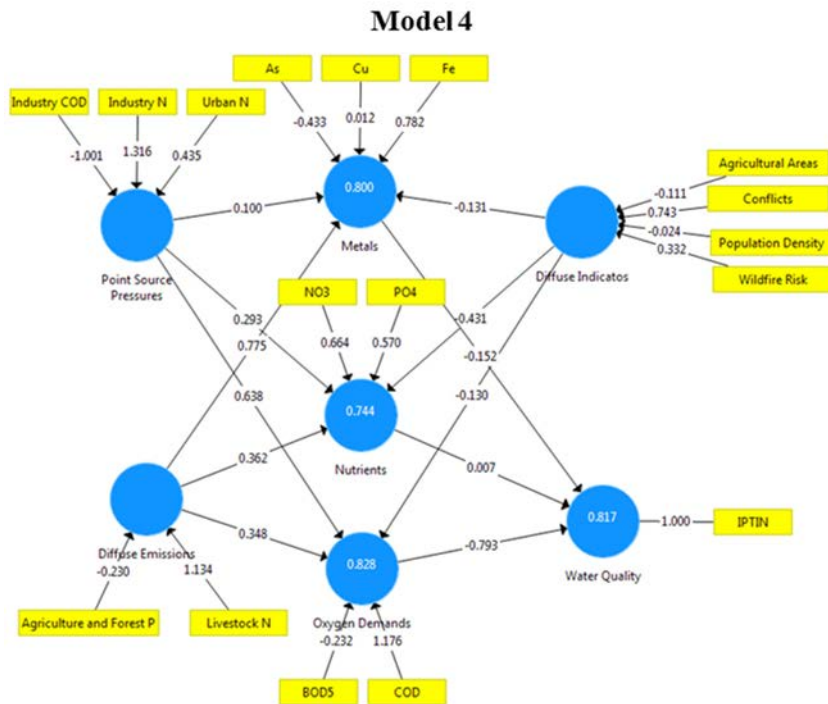Figure 4: Structural equation models 1, 2 and 3.

Figure 5: Model 4.

Most of the path coefficients from all the models seem to have a correct sign. For example in model 1 "pollution sources" increase "water contamination" (pc = 0.937) while "water contamination" decreases "water quality" (pc = -0.965). But in other models, there are some path coefficients that do not have an expected value. In model 2 "nutrients" have a null effect in "water quality" this can be possible if the other pollution sources have a stronger impact in the loss of ecological integrity, so the effect of "nutrients" is absorbed by other variables. In model 3 "point source pressures" and "diffuse emissions" increase "water contamination", but "diffuse indicators" decrease it (pc = -0.211). Possibly this happens because the other LVs have a direct impact in "water contamination", while "diffuse indicators" are not so directly associated to "water contamination" and is attributed a negative sign because the measured variables that composed it are positively correlated with "water contamination". For model 4 is clearly verified the same, besides both "point source pressures" and "diffuse emissions" have a positive pc in "nutrients", "oxygen demands" and "metals", but "diffuse indicators" path coefficients are negative. In model 4 is still noticed that nutrients have an almost null effect in "water quality" just like in model 2.

To compare the effect of each latent variable and measured variable among the four models it was calculated the product between the total effect of each LV in "Water Quality" with the weight of each measured variable that composes it (Table 2). For instance, in model 1 the direct effect (or path coefficient) of "pollution sources" in "water quality" is 0 because these variables are not connected, but since "pollution sources" composes "water contamination" (pc = 0.937) that composes "water quality" (pc = -0.965), the indirect effect of "pollution sources" in "water quality" is the product of (-0.965 × 0.937) which is -0.904.

The total effect is the sum of the direct and indirect effects, which is still -0.904 in this case. In model 1 the product of the total effect with the weight of nitrogen from urban discharges (Urban N weight = 0.159) is -0.144. This procedure was applied to all variables in all models, as is shown in Table 2.

Table 2: Product of total effect on water quality and the weight for each measured variable.

	Model 1	Model 2	Model 3	Model 4
BOD_5	0.181	0.164	0.164	0.184
COD	-0.832	-0.996	-0.807	-0.933
NO_3	-0.186	0.000	-0.161	0.005
PO_4	-0.260	0.000	-0.276	0.004
As	0.300	0.073	0.303	0.066
Cu	0.102	-0.005	0.104	-0.002
Fe	0.236	-0.120	0.214	-0.119
Agricultural Areas	-0.054	-0.067	0.021	-0.013
Agriculture and Forest P	0.073	0.082	0.105	0.090
Conflicts	0.135	0.221	0.106	0.089
Industry COD	0.783	0.356	0.751	0.520
Industry N	-0.925	-0.504	-0.888	-0.683
Livestock N	-0.297	-0.592	-0.306	-0.444
Population Density	-0.156	-0.110	-0.018	-0.003
Urban N	-0.144	-0.203	-0.175	-0.226
Wildfire Risk	-0.033	-0.075	0.129	0.040

By analysing the reported results in Table 2, it can be assured that chemical oxygen demand (COD) is the surface water parameter that has the most substantial effect in the decrease of "water quality", while biological oxygen demand (BOD_5) do not. This difference might be due to the fact that COD in surface water is high when many industrial effluents are discharged due to the low concentration of organic matter. In the other hand, BOD high values are typical from rivers with urban effluents due to the high organic loads. Probably the model attributed a stronger and negative effect to COD, and weaker and positive to BOD since industrial effluent discharge has a higher impact in the river basin, rather than urban effluents. Arsenic (As) concentration is another parameter that has a similar effect in all models, and it does not low "water quality". This cannot be viewed that As increases water quality, because the nefarious effects of arsenic in the environment are well-known, but according to the applied modulation technique this variable is not restraining water quality in Ave river basin, probably because pollution sources that release arsenic few, which is concordant to the decreasing tendency of heavy metal pollution [20].

According to the models, the contribution of NO3, PO4, Fe and Cu in $IPtI_N$ is inconclusive. Since is seen that in models 1 and 3 nutrients decrease water quality, while in models 2 and four, nutrients do not increase water quality, but Cu and Fe decrease. The main difference from models 1 and 3 to models 2 and 4 is that the latent variable "contamination" is substituted by three other latent variables, where the type of contamination becomes differentiated into "metals", "nutrients" and "oxygen demands". In models 2 and 4 is possible to trace the individual effects of "metals", "nutrients" and "oxygen demands". Maybe the effects of nutrients occur when these variables are gathered with other surface water parameters, which can be called as a combined effect. While for Fe and Cu, the effect might occur isolated. Anyhow for both of these variables, their contribution might be

underestimated in different models due to the fact that COD has high influence in water quality and might absorb their effect. As it can be seen in supplementary material, COD is the variable with the strongest correlation with $IPtI_N$ ($r_{Pearson}$ = -0.888), and so is the surface water parameter with strongest decrease effect with $IPtI_N$ along with the four models.

For agricultural areas, the product is positive for models 1, 2 and 4 but near to zero for all, so this variable does not seem to restrain water quality. For wildfire risk in models 1,2 and 4 the product is near to zero, only in model 3 the product is positive and considerable, so in an overall perspective, the effect of this variable is unknown for Ave river basin water quality. The releases of phosphorous from forest and agriculture, land use conflicts and released COD from industry are pressures that do not restrain water quality. Discharges of nitrogen from industry, livestock, and urban discharges are the pressures that considerably decrease water quality in the river basin, while for population density it is verified that besides the sign of the product is negative for all models, the results are near to zero, so it is assumed as a minor pressure.

The SEM-PLS prediction of $IPtI_N$ for each sub-basin were interpolated into raster files, for each watercourse. It was calculated the average value and then classified as very poor, poor, moderate, good and excellent according to the $IPtI_N$ Portuguese classification [25], Fig. 6.

Since adjusted R squared values of "water quality" are considerably high, the prediction maps of are quite similar measured $IPtI_N$. The classification of water courses majorly matches for all models, the unmatched watercourses are presented in Fig. 6 (F). For a clear distinction of the prediction results, the percentage of Ave River basin watercourses length is shown in Table 3, comparing the measured classification with the predicted.
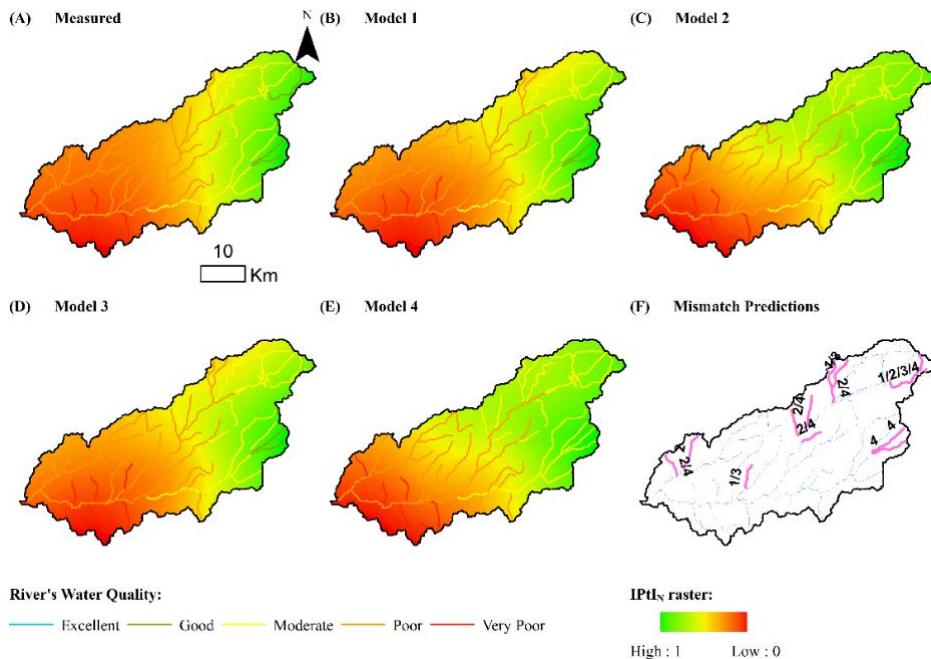

Figure 6: $IPtI_N$ comparison: (a) Measured; (b) Model 1 prediction; (c) Model 2 prediction; (d) Model 3 prediction; (e) Model 4 prediction; and (f) Mismatch predictions.

Table 3: Comparison of measured IPtI$_N$ with prediction for each watercourse.

Measured → Model	Model 1	Model 2	Model 3	Model 4
Very Poor → Very Poor	7.9%	7.9%	7.9%	7.9%
Poor → Very Poor	1.3%	3.1%	1.3%	1.3%
Poor → Poor	53.2%	44.9%	53.2%	46.8%
Poor → Moderate	0.0%	6.4%	0.0%	6.4%
Moderate → Poor	1.6%	0.0%	1.6%	0.0%
Moderate → Moderate	29.3%	30.9%	29.3%	30.9%
Good → Moderate	3.0%	3.0%	3.0%	6.8%
Good → Good	3.7%	3.7%	3.7%	0.0%

For models 1 and 3, 94.1% of the prediction matches, while for models 2 and 4, 81.6% and 81.7% respectively. Models 1 and 3 have exactly the same prediction classification and the highest percentage of classification matches. These models have in common that the surface water parameters are all gathered in the same LV, which resulted in higher predictions, while in models 2 and 4 lower matching percentages were obtained. The used algorithm in SEM-PLS models tends to maximize all the R-Squared Values for all latent variables, so as more latent variables are placed in the model the R squared value of crucial variable "water quality" decreases, so since models 1 and 3 have only 2 latent variables that have a calculated R-squared, they can explain more variation of IPtI$_N$, which make them more adequate for prediction purposes rather than models 2 and 4.

According to the Portuguese legislation (Decree-law no. 236/98), the maximum allowed concentration for drinking purposes are 250 mg/l (NO_3), 50 µg/l (As), 100 µg/l (Cu), and 50 µg/l (Fe). In any the locations of Ave River basin, these thresholds were not overlapped. For COD, BOD and PO_4 the limit values are established for captured water for the production of human consumption water, respectively 3 mg/l, 30 mg/l and 400 µg/l. For the 92 analysed river basins it was founded that the concentration of BOD and PO_4 exceeded the limit values in 1 and 12 sub-basins, respectively. Besides the number of exceedances is not high, it is essential to refer that even when the concentrations are below legal limits there is still possible to track damages in ecosystems [29]. So when water quality is studied is crucial to use bioindicators, because not only chemical status determines water quality, but also ecological integrity.

SEM-PLS was the chosen prediction technique, due to the fact that the presented models are simple to analyse. The cause-effect relationships were clearly demonstrated in four models and the mismatch predictions were low. When modelling techniques are used, is important to have not only a robust and explicative model but also, when possible, simple to analyse, by readers and or users, because models are communications tools [30].

4 CONCLUSIONS

Four SEM-PLS data models were created with the same dataset, this study can lead to conclude that each model has its purpose. Model 1 is the simplest model, as an advantage is the most accessible model to interpret for practically any reader and like model 3, they are the most suitable for prediction purposes. Models 2 and 3 differentiate types of contamination and pollution sources separately, which makes them quite useful when these subjects are studied disjointedly. Model 4 contains more information than all the other models, due to the higher number of latent variables it can lead to a better understanding of the relationship between different pollution sources and released contaminants.

In further studies is intended to use these models to predict $IPtI_N$ for Ave River basin, by using new pressure values in sub-basins, in order to verify the validity of these models for future data. As it was demonstrated in another study [17] that the application of SEM-PLS in unpolluted basins can result in lower R-Squared values what would lead to an incorrect classification of $IPtI_N$. Anyhow this methodology should be compared and tested with another river basins, but it is warned that not the same variables can be always used because the dismissed variables through multicollinearity might not be the same.

ACKNOWLEDGEMENTS

REFERENCES

[1]   Loucks, D.P., van Beek, E., Stedinger, J.R., Dijkman, J.P.M. & Villars, M.T., Water quality modelling and prediction. *Water Resources Systems Planning and Management: An Introduction to Methods, Models and Applications*, eds D.P Loucks & E. van Beek, UNESCO Publishing, 2005.

[2]   Thakur, A.K., Model: Mechanistic vs. empirical. *New Trends in Pharmacokinetics*, eds A. Rescigno & A.K. Thakur, Springer, 2012.

[3]   Streeter, H.W. & Phelps, E.B., A study of the pollution and natural purification of the Ohio River. Public Health Bulletin (Washington DC), 1925.

[4]   Neitsch, S., Arnold, J., Kiniry, J. & Williams, J., Soil and water assessment tool theoretical documentation version 2009. Texas Water Resource Institute, TR-406, 2009.

[5]   Wool, T., Ambrose, R., Martin, J., Comer, E.A. & Tech, T., Water quality analysis simulation program (WASP), User's Manual, Version, 2006.

[6]   MIKE 11, *Encyclopedia of Hydrological Sciences*, 2006.

[7]   Donigian, J.A.S., Bicknell, B.R., Imhoff, J.C. & Singh, V.P., Hydrological simulation program – Fortran (HSPF), *Computer Models of Watershed Hydrology*, ed. V.P. Singh, Water Resources Publications: Colorado, 1995.

[8]   Leon, L. & Smith, R., Lake Erie hypoxia simulations with ELCOM-CAEDYM, 3rd Biennial Meeting of the International Environmental Modelling and Software Society, 2006.

[9]   Binz Astrachan, C., Patelb, V. & Wanzenried, G., A comparative study of CB-SEM and PLS-SEM for theory development in family firm research. *Journal of Family Business Strategy*, **5**(1), pp. 116–128, 2014.

[10]   Hair, J.F.M., Hult, G.T., Ringle, C. & Sarstedt, M., *A Primer on Partial Least Squares Structural Equation Modeling*, SAGE Publishing, 2014.

[11]   Zou, S. & Yu, Y.-S., A general structural equation model for river water quality data. *Journal of Hydrology*, **162**(1–2), pp. 197–209, 1994.

[12] Chenini, I. & Khemiri, S., Evaluation of ground water quality using multiple linear regression and structural equation modeling. *International Journal of Environmental Science & Technology*, **6**(3), pp. 509–519, 2009.

[13] Wu, E.M.-Y., Tsai, C.C., Cheng, J.F., Kuo, S.L. & Lu, W.T., The application of water quality monitoring data in a reservoir watershed using AMOS confirmatory factor analyses. *Environmental Modeling & Assessment*, **19**(4), pp. 325–333, 2014.

[14] Levêque J.G. & Burns, R.C., A structural equation modeling approach to water quality perceptions. *Journal of Environmental Management*, **197**, pp. 440–447, 2017.

[15] Nugroho, A.R. & Masduqi, A., Structural equation modelling as instrument for water pollution factor analysis. *Proceedings of the 4th International Conference Adv. Appl. Sci. Environ. Technol.*, pp. 28–32, 2016.

[16] Sanches Fernandes, L.F., Fernandes, A.C.P., Ferreira, A.R.L., Cortes, R.M.V. & Pacheco, F.A.L., A partial least squares path modeling analysis for the understanding of biodiversity loss in rural and urban watersheds in Portugal. *Science of the Total Environment*, **626**, pp. 1069–1085, 2018.

[17] Fernandes, A., Ferreira, A., Fernandes, L.S., Cortes, R. & Pacheco, F., Path modelling analysis of pollution sources and environmental consequences in river basins. *WIT Transactions on Ecology and the Environment*, vol. 228, WIT Press: Southampton and Boston, pp. 79–87, 2018.

[18] Gonçalves, E.P.R., Boaventura, R.A.R. & Mouvet, C., Sediments and aquatic mosses as pollution indicators for heavy metals in the Ave river basin (Portugal). *Science of the Total Environment*, **114**, pp. 7–24, 1992.

[19] Soares, H.M.V.M., Boaventura, R.A.R., Machado, A.A.S.C. & Esteves Da Silva, J.C.G., Sediments as monitors of heavy metal contamination in the Ave river basin (Portugal): Multivariate analysis of data. *Environmental Pollution*, **105**(3), pp. 311–323, 1999.

[20] Alves, C., Boaventura, R. & Soares, H., Evaluation of heavy metals pollution loadings in the sediments of the Ave River basin (Portugal). *Soil and Sediment Contamination: An International Journal*, **18**, pp. 603–618, 2009.

[21] Dunck, B., Lima-Fernandes, E., Cássio, F., Cunha, A., Rodrigues, L. & Pascoal, C., Responses of primary production, leaf litter decomposition and associated communities to stream eutrophication. *Environmental Pollution*, **202**, pp. 32–40, 2015.

[22] Ramos, S., Amorim, E., Elliott, M., Cabral, H. & Bordalo, A.A., Early life stages of fishes as indicators of estuarine ecosystem health. *Ecological Indicators*, **19**, pp. 172–183, 2012.

[23] Ribeiro, C.M.R. et al., Anthropogenic pressure in a Portuguese river: Endocrine-disrupting compounds, trace elements and nutrients. *Journal of Environmental Science and Health – Part A Toxic/Hazardous Substances and Environmental Engineering*, **51**(12), pp. 1043–1052, 2016.

[24] Pacheco, F.A.L., Varandas, S.G.P., Sanches Fernandes, L.F. & Valle Junior, R.F., Soil losses in rural watersheds with environmental land use conflicts. *Science of the Total Environment*, **485–486**, pp. 110–120, 2014.

[25] INAG, *Critérios para a Classificação do Estado das Massas de Água Superficiais – Rios e Albufeiras*, 2009.

[26] ESRI, ArcMap (Version 10). New York St. Redlands, USA, 2010.

[27] ESRI, ArcHydro Tools for ArcGIS 10 – Tutorial, 2012.

[28] Monecke A. & Leisch, F., semPLS: Structural equation modeling using partial least squares. *Journal of Statistical Software*, **48**(3), 2012.

[29] Fonseca, A.R., Sanches Fernandes, L.F., Fontainhas-Fernandes, A., Monteiro, S.M. & Pacheco, F.A.L., From catchment to fish: Impact of anthropogenic pressures on gill histopathology. *Science of the Total Environment*, **550**, pp. 972–986, 2016.

[30] Caro, J.J., Briggs, A.H., Siebert, U. & Kuntz, K.M., Modeling good research practices-overview: A report of the ISPOR-SMDM modeling good research practices task force-1. *Medical Decision Making*, **32**(5), pp. 667–677, 2012.