Accepted Manuscript

Water security and watershed management assessed through the modelling of hydrology and ecological integrity: a study in the Galicia-Costa (NW Spain)


DOI: https://doi.org/10.1016/j.scitotenv.2020.143905

To appear in: Science of The Total Environment

Peer reviewed version

Please cite this article as:


General rights:

© 2020 Elsevier B.V. This article is distributed under the terms and conditions of the Creative Commons Attribution-Noncommercial-No Derivatives (CC BY-NC-ND) licenses https://creativecommons.org/licenses/by-nc-nd/4.0/
Water security and watershed management assessed through the modelling of hydrology and ecological integrity: a study in the Galicia-Costa (NW Spain)

Carolina Acuña-Alonso 1*, António Carlos Pinheiro Fernandes 3, Xana Álvarez 1, Enrique Valero 1, Fernando António Leal Pacheco 2, Simone Da Graça Pinto Varandas 3, Daniela Patrícia Salgado Terêncio 2,3, Luís Filipe Sanches Fernandes 3

1 Forestry Engineering School, University of Vigo, Campus A Xunqueira s/n., 36005, Pontevedra, Spain. e-mail: carolina.alonso@uvigo.es (C. Acuña-Alonso); xaalvarez@uvigo.es (X. Álvarez); evalero@uvigo.es (E. Valero)

2 Center of Chemistry of Vila Real, University of Trás-os-Montes e Alto Douro, Ap. 1013, 5001-801 Vila Real, Portugal. e-mail: fpacheco@utad.pt (F.A.L. Pacheco)

3 Center for Research and Agro-environmental and Biological Technologies, University of Trás-os-Montes e Alto Douro, Ap. 1013, 5001-801 Vila Real, Portugal. e-mail: simonev@utad.pt (S.D.G. Varandas); acpf91@utad.pt (A.C.P. Fernandes); dterencio@utad.pt (D.P.S. Terêncio); lfilipe@utad.pt (L.F.S. Fernandes)

* Correspondence: carolina.alonso@uvigo.es (C. Acuña-Alonso)
**Highlights**

Partial Least Squares Path Modeling can be used to assess the water quality

Pressures and water contamination describe biodiversity loss in Galicia-Costa

The variables "Anthropogenic" and “SWP” decrease “Ecological Integrity”

Land use planning is crucial for water quality

The deterioration of riparian vegetation damages the good state of water security
ABSTRACT

Water management is a crucial tool for addressing the increasing uncertainties caused by climate change, biodiversity loss and the conditions of socioeconomic limits. The multiple factors affecting water resources need to be successfully managed to achieve optimal governance and thus move towards water security. This study seeks to obtain a holistic vision of the various threats that affect the ecological integrity of the basins that form the hydrological district of Galicia-Costa, through the method of partial least squares path modelling (PLS-PM). The data is analysed overall for the hydrological years from 2009 to 2015. The independent latent variables are “Anthropogenic” (comprising the percentage of water bodies with edges alongside artificial surfaces, the percentage connected to artificial land use patches, the edge density of artificial surfaces and population density) and “Nature” (edge density of forestry land uses, edge length of land water bodies alongside forested areas and the percentage of land occupied by the largest patch of forest). The dependent latent variables are “SWP”, which represents surface water parameters (biological oxygen demand, chlorides, conductivity and dissolved iron) and “Ecological Integrity” (METI Bioindicator). From an overall perspective, the PLS-PM results reveal that 69.0% of “SWP” is predicted by the independent variables ($R^2 = 0.690$), “Anthropogenic” contributes by increasing SWP ($\beta = 0.471$), while “Nature” decreases the concentration of SWP ($\beta = -0.523$), which indicates the polluting parameters in the water. The variables "Anthropogenic" ($\beta = -0.351$) and “SWP” ($\beta = -265$) lower the quality of “Ecological Integrity”. This variable must be managed through soil conservation measures for the benefit of water security. This study has been able to identify and quantify the variables that increase contaminant concentration and decrease ecological integrity, providing a promising methodology that facilitates protection and correction measures to guarantee water safety.
1. Introduction

Water management is one of the main current global challenges due to the role played by fresh water as a resource for sustaining life and human needs (Sappa et al., 2019). Moreover, with social and demographic expansion, many threats to water quality have arisen and reached worrying levels around the globe (A. C. Fernandes et al., 2019b). Kummu et al., (2016) find that while water consumption has increased fourfold in the past 100 years, the population likely to be affected by water scarcity has increased from 0.24 billion to 3.8 billion people. Land uses also have direct impacts on hydrologic systems within watersheds (Bolstad and Swank, 1997), in addition to climate and soil conditions (Tong and Chen, 2002). The different types of land use are closely related to the characteristics of human activities, which in turn determine the anthropogenic substances carried into hydrologic systems through drainage or runoff processes (Lee et al., 2009). The World Health Organisation (WHO) (Worl Health Organization, 2005) has reported that changes in land use have been the main driver of the deterioration of freshwater ecosystems in the last 50 to 100 years in addition to the loss of global biodiversity. Furthermore, single and/or multi-purpose water infrastructures have transformed freshwater ecosystems, and only a third of all rivers longer than 1000 km now flow freely to the ocean (Grantham et al., 2019) (Grill et al., 2019). Therefore, it is necessary to understand the consequences for ecological integrity of a wide range of changes in the physical habitat. In this sense, the Water Framework Directive (European Commission, 2000) is the main legislative driver in specifying that hydromorphology should underpin "good ecological status". Moreover, various studies have analysed and quantified the role of pollutants in basins (Marrugo-Negrete et al., 2017) (Yang et al., 2016). Effluent discharges have the potential to significantly alter many different aspects of aquatic systems, including nutrient uptake efficiency (Haggard et al., 2001). All these alterations, both physical and chemical, cause impacts on biodiversity, ecological structure and function or use of water (Levin et al., 2019) (Jackson et al., 2016). Thus, water quality should be assessed via physical, chemical and biological attributes in order to provide a complete spectrum of
information for appropriate water governance (Iliopoulou-Georgudaki et al., 2003). Accordingly, bioindicators are used to examine ecological integrity (Anyanwu et al., 2019). There are several indices whose objective is to analyse and quantify the effect of pressures on fresh water resources, to determine whether they are excessive or damage ecological integrity. Examples include the IBMWP (Iberian Biological Monitoring Working Party) index (Alba-Tercedor and Sánchez-Ortega, 1978) (Alba Tercedor et al., 2002), the TDIL (Trophic Diatom Index for Lakes) (Stenger-Kovács et al., 2007) and the B-IBI (the index of biotic integrity) (Weisberg et al., 1997).

Despite great experience in European countries with the use of these indices, the Water Framework Directive (WFD), which came into force in 2000, has necessitated a revision of the indices and methods in use or the creation of new, more accurate methods to present greater adaptability to changes in aquatic ecosystems (Munné and Prat, 2009). Spain has developed the METI (Type Specific Multimetric Index) (MAGRAMA, 2015), which is used to assess the ecological status of north-western rivers in compliance with the Water Framework Directive (European Commission, 2000).

Integrated water resource management is a well-established framework in the water sector that has been adopted by governments in all regions and at all levels of economic development (Jensen and Nair, 2019). Its main goals include strengthening water security, increasing sustainability and ensuring equitable access to services. The cost of global water insecurity is estimated to be $500 billion per annum and is likely to be a drag on the world economy of 1% or more of gross domestic product (GDP) (Wiberg, 2015). The concept of water security has received increased attention over the past decade. It has multiple definitions depending on which discipline studies it. From the point of view of environmental sciences, water security seeks to: (1) guarantee access to water functions and services for humans and the environment; (2) ensure the availability of water in terms of quality and quantity; and (3) minimise the impacts of hydrological variability (Cook and Bakker, 2012). The Global Water Partnership (Global Water Partnership, 2000) has defined water security as ‘access to adequate safe water at an affordable
cost, ensuring that the natural environment is protected and enhanced; thus incorporating ecological dimensions’. More recently, the concept of water security has emphasised the water-energy-food (WEF) nexus: the complex interactions within the wider framework of water, food and energy which mean that actions in each sector can have effects in the others, and that water security, energy security and food security are closely linked (FAO, 2014). The WEF nexus is increasingly perceived as a promising approach for overcoming governance failures in dealing with complex and interconnected resource management challenges (Pahl-Wostl, 2019a). Furthermore, this nexus is increasingly recognised as a conceptual framework capable of supporting the efficient implementation of the Sustainable Development Goals (SDGs) (Terrapon-Pfaff et al., 2018; Yillia, 2016). Water security has become one of the main objectives of environmental governance and resource management (Gerlak et al., 2018). It is therefore key to planetary resilience, mitigating the effects of climate change (Keys et al., 2019). As a result, ensuring water governance has become a topic of great scientific and political interest (Pahl-Wostl, 2019b). It is therefore necessary to implement water governance policies which, from an ecological perspective, enable the focus to be placed on local knowledge of the use of resources and access to resources (Xie et al., 2019). The World Economic Forum (Terrapon-Pfaff et al., 2018) concluded that improved water governance was necessary “to adapt to climate change and accommodate a growing population and economic development”.

To guarantee water security, it is necessary to analyse the sources of pollution (A. C. Fernandes et al., 2019b) and quantify those variables that influence degradation most. To that end, statistical models can be useful (Pacheco and Landim, 2005). These models can be based on eigenvectors or structural equation models (SEM). CB-SEM (Covariance-based SEM) and PLS-PM (PLS-Path Modelling) are the most common (Astrachan et al., 2014). In the former, the estimation procedure is based on a maximum likelihood estimation, while in PLS-PM a nonparametric partial-least squares path modelling (Tubadji and Nijkamp, 2015), is based on ordinary least squares regression. The benefits of each type of SEM differ depending on the
study, purpose and dataset. PLS-PM originated in social sciences (Wold, 1966), and for decades it has been widely and dominantly used in the field of Information Systems (IS) (Benitez-Amado et al., 2017). Subsequent applications have expanded the use of PLS-PM to other areas including the environmental sciences (Fernandes et al., 2019b). In this study, PLS-PM was applied in the hydrological district of Galicia-Costa (the coast of the Galicia region of north-western Spain). The study seeks to take a step forward in multivariate statistical modelling by using river basins, with the integration of PLS-PM in an environmental scope. This analysis provides additional insights about the dataset’s structure, helping gain an understanding of direct and indirect interactions between numerous latent factors (Astrachan et al., 2014). The main objective of this manuscript is to determine, analyse and quantify cause-effect interactions between pollution sources, land uses, surface water parameters and ecological integrity for different hydrological years and overall across those years by using the PLS-PM model. The manuscript is presented as an innovative analysis, since there are few environmental studies that use SEM models. Moreover, those few apply such models for the same basin (Sanches Fernandes et al., 2018), (Fernandes et al., 2019), (Fernandes et al., 2019a), while here PLS-PM is applied to 18 basins with a total of 40 sampling sites, providing a large amount of information. Analysing this group of basins belonging to the Galicia-Costa Hydrographic Demarcation (making up a management unit) favours the integrated management of water resources. In this way, an innovative analysis is presented to support decision-making by the corresponding management unit, prioritising optimal water governance and strengthening water security.

2. Methodology

2.1. Study Area

The hydrological district of Galicia-Costa is located in the Galicia region of north-western Spain (Figure 1). This district is made up of 19 sub-basins associated with medium/large rivers in the
interior of the region. The coastal basins are smaller, pour directly into the sea and have greater torrentiality. The district extends over an area of 12,991 km² (44.05 % of the territory of Galicia) (Augas de Galicia, 2015). According to data from the Galician Institute of Statistics (IGE), the population of the Galicia-Costa hydrographic district in 2013 was 2,070,645 (Instituto Galego de Estatistica (IGE, 2015). Galicia has a highly scattered population (about 40% live in small rural villages with fewer than 500 inhabitants or in isolated houses) (Raposo et al., 2013). This characteristic trait of rural Galicia and its complicated geography entail an overcost in the provision of public services (Macho, 2013).

Figure 1. Study area: watersheds analysed.

Using the Information System on Land Use in Spain (SIOSE) usage map but with the generic CORINE Land Cover code, the 33 land uses are reclassified into generic uses referred to as "Artificial surfaces", "Agricultural areas", "Forest and semi natural areas" and "Water bodies". According to SIOSE for 2006, the breakdown for the hydrological district of Galicia-Costa is forest and semi-natural areas (61.44%), agricultural areas (34.41%), artificial surfaces (4.41 %) and
water bodies (0.49%) (Figure 2). In recent years, the forest area has increased gradually and the availability of agricultural land has decreased (Ferreiro-Domínguez et al., 2017) due to land abandonment. In the methodology applied in this studied, the part of the basin upstream from each measurement point is used, so it is assumed that all pollutant transfer phenomena only influence downstream. That is why the basins designed in this study comprise only the drainage area from pollution sources to the measurement point. The soil of Galicia-Costa is mainly characterised by two dominant types: regosols (which cover 57.6% of the total area) and leptosols (which cover 40.0% of the total area), and to a lesser extent by histosols and podzoles, which account for only 2.4% of the total area (FAO, 1990).

Figure 2. Land use and population density in study area.

According to the Köppen-Geiger climate classification (Kottek et al., 2006), almost the whole of Galicia has a warm temperate climate with dry, warm summers (Warm-summer Mediterranean climate, Csb), characterised by maximum seasonal rainfall in winter and minimum in summer (Raposo et al., 2013). By contrast, the Cantabrian area has a warm temperate climate which is
fully humid with warm summers (Temperate oceanic climate, Cfb). Based on the 1940/1941-2011/2012 series, the average precipitation is 20,039 hm³/year. Rainfall varies widely, with a maximum mean of 3,460 mm/year and a minimum mean of 248 mm/year, making an average of 1,543 mm/year.

Galicia-Costa is characterised by high availability of groundwater due to its shallow water table. Groundwater is thus an optimal source of water supply in rural areas (Raposo et al., 2012). Both the aquifers and the population are widely dispersed. 75% of the registered water sources use groundwater, but surface water represents the largest resource in terms of volume consumed (Xunta Galicia, 2005). These water facilities are mainly shallow wells or springs, with little storage capacity and short residence periods, making them highly vulnerable to climate change (Raposo et al., 2013).

2.2. Conceptual framework model

For the present study, different interactions between environmental data from the Galicia region were exposed using PLS-PM models for different hydrological years. Data were gathered on surface water parameters between 2010 and 2015. This information was provided by Augas de Galicia (https://augasdegalicia.xunta.gal/). Only eighteen of the nineteen basins that make up the Galicia-Costa Demarcation were analysed, as one lacked the data required for the analysis. The original forty physicochemical parameters were narrowed down to a group of four (BOD₅, chlorides, conductivity and dissolved iron), which were associated with a decrease in the METI biological index in the PLS-PM models. The METI index was used to assess the biological status of aquatic ecosystems which depended on the macroinvertebrate community composition. High scores indicate low pollution and vice versa (MAGRAMA, 2015). The average figure per hydrological year (1 October to 30 September) was calculated for all the surface water physicochemical parameters and METI metrics. Land use and ground cover maps were created using an image from the Information System on Land Use in Spain (Gobierno de España, 2016)
with 12.5m spatial resolution (https://www.siose.es/). For each drainage section, landscape metrics were calculated for generic land uses (Appendix 1), water bodies, artificial, forestry and agricultural areas using a python toolbox embedded in ArcMAP (Adamczyk and Tiede, 2017) and the 6 that best suited the model were selected (Appendix 1). The digital elevation model was obtained with a resolution of 25 per 25m (http://www.eea.europa.eu) for the 40 sampling sites, the respective upstream areas for each point were outlined using ArcMap (ESRI, 2012) and ArcHydro tools (Maidment and Morehouse, 2002). The population density for each drainage section was also estimated. First, a shapefile was created containing the total population of each municipality, divided into the respective areas. Then the shapefile was converted into a raster file containing the population density of each municipality and was used in the Zonal statistics as a table ArcMap tool (ESRI, 2012) to extract the average population density of each drainage area.

2.3. Partial least squares – Path modelling

The Partial Least Squares (PLS) method was originally developed by Wold (Wold, 1980, 1966). In this method, causal paths are established between blocks of variables called latent variables (LV) or constructs. In Partial Least Squares-Path Modelling, the LV and paths form the inner (also called structural) model, while the measured variables (MV) represent the outer or external model (Wold, 1980). In the inner model, the connections between LV are quantified through path coefficients (β), while the links between LV and MV in the outer model are quantified through weights (W) (Hair Jr. et al., 2016).
Figure 3. Diagram of complete PLS-PM models for Galicia-Costa. The circles represent the latent variables, namely “Anthropogenic”, “Nature”, “SPW” and “Ecological integrity”. The rectangles represent formation variables (see description in Table 1). The arrows represent the links between formation and associated latent variables, and between related latent variables, while arrow labels are weights and path coefficients that quantify those links.

Table 1. List of measured variables used as source data for Partial Least Squares–Path Modeling (PLS-PM). Besides identifying and describing variables and giving the units in which they are measured, the table gives indications on usage in the PLS-PM models and on the data sources.
With all the data gathered, a group of representative variables was selected to create formative Partial Least Squares-Path Models (PLS-PM). Seven models were built: six to represent the hydrological years between 2010-2015 and one using average values across the hydrological years, to cap the global interactions over the periods studied, as shown in Figure 3. For the models, the input data for surface water parameters and METI were different in each hydrological year, while the input data for variables were the same for each one. The data were grouped into 4 latent variables: “Ecological Integrity” (containing METI), “SWP” (which represents surface water parameters such as BOD$_5$, chlorides, conductivity and dissolved iron), “Nature” (containing landscape metrics of forestry land uses such as edge density of forestry land uses, edge length of inland water bodies alongside forested areas and the percentage of land occupied by the largest patch of forest) and “Anthropogenic” (containing the population

<table>
<thead>
<tr>
<th>Latent variables</th>
<th>Measured Variable</th>
<th>Units</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BOD$_5$</td>
<td>mg/l O$_2$</td>
<td>Biological oxygen demand</td>
</tr>
<tr>
<td></td>
<td>Chlorides</td>
<td>mg/l Cl</td>
<td>Chlorides</td>
</tr>
<tr>
<td></td>
<td>Conductivity</td>
<td>µS/cm</td>
<td>Conductivity</td>
</tr>
<tr>
<td></td>
<td>Dissolved iron</td>
<td>mg/l</td>
<td>Dissolved iron</td>
</tr>
<tr>
<td></td>
<td>ed_ (FOR)</td>
<td></td>
<td>Edge density of forestry land uses</td>
</tr>
<tr>
<td></td>
<td>el_ (WAB) with_ (FOR)</td>
<td></td>
<td>Edge length of land water bodies alongside forested areas</td>
</tr>
<tr>
<td></td>
<td>lpi_ (FOR)</td>
<td></td>
<td>Percentage of land occupied by the largest patch of forest</td>
</tr>
<tr>
<td></td>
<td>cce(ART) with_ (WAB)</td>
<td></td>
<td>Percentage of artificial land uses beside water bodies</td>
</tr>
<tr>
<td></td>
<td>ci_pp_ (ART)</td>
<td></td>
<td>Percentage of connected of artificial land uses</td>
</tr>
<tr>
<td></td>
<td>ed_ (ART)</td>
<td></td>
<td>Edge density of artificial surfaces</td>
</tr>
<tr>
<td></td>
<td>Population density</td>
<td></td>
<td>Population density</td>
</tr>
<tr>
<td></td>
<td>METI</td>
<td></td>
<td>METI Bioindicator</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Latent variables</th>
<th>Measured Variable</th>
<th>Units</th>
<th>Description</th>
</tr>
</thead>
</table>
density and metrics for artificial land uses, e.g. the percentage of artificial land uses beside water bodies, the percentage of connection with artificial land uses and the edge density of artificial surfaces), as indicators of pollution sources. The models were created to analyse the cause-effect interactions between pollution sources, natural land uses, (SWP), and “Ecological Integrity”. The latent variables selected were created and connected with scientific consistency, inspired by other PLS-PM models (Oliveira et al., 2019)(Sanches Fernandes et al., 2018). The latent variable “Anthropogenic” represents the threats to ecological integrity, while “Nature” represents phenomena that improve water quality. Both these latent variables are connected to “SWP” and “Ecological Integrity”, which corresponds to METI scores, since “Nature” and “Anthropogenic” affect the concentration of SWP and “Ecological Integrity” directly. The links between the latent variables are called the “inner model”. For the present study the inner model comprises 2 exogenous variables (not dependent on any other latent variable) – “Nature” and “Anthropogenic” – and 2 exogenous variables (composed by other latent variables): “SWP” and “Ecological Integrity” (Table 1).

According to Moran et al., (2018), Hoyle (1999) and Garson (2016), the minimum sample size can be defined as ten times the maximum number of measured variables that form a latent variable. Taking into account that this study contains a total of forty sampling sites/drainage areas, the maximum of the measured variables is four. In line with this constraint, various models with different combinations of measured variables were tested. A group of variables was chosen in which most cause-effect interactions were largely concordant with the theory and the resulting variance inflation factors (VIFs) were below five (Kock, 2015). In this study, seven PLS-PM models were created with the same structure as the one represented in Figure 3. These models were formed with a total of twelve measured variables (yellow boxes) and four latent variables (coloured circles). After the models were constructed, the PLS-PM algorithm was applied. Through an iterative process carried out by the Smart-PLS program, weights for each measured variable (connections between yellow boxes and circles) were established, and path
coefficients (also called direct effects) for each connection between latent variables to maximise the R² values for each endogenous latent variable, indicating how much variance in the measured score is explained by the predicted score (Figure 3).

In PLS-PM models two types of score can be established for each latent variable, measured score and predicted score. For exogenous latent variables only the measured score is established, which is the sum of the product of the measured variable by the respective weight. For endogenous latent variables, a predicted score is also set which is the sum of the product of each latent variable that composes it by the respective path coefficient (direct effect). For this reason, only endogenous latent variables have an associated R², which is the variance in the measured score which is explained by the predicted score.

3. Results and discussion

3.1. Partial Least Squares-Path Modelling (overall)

The starting point for this discussion is the observation that the models portray the combination of “SWP” better than “Ecological Integrity”, since the R² of “SWP” is always higher than that of “Ecological Integrity”. For example, in the “Overall” model, Figure 4, the resulting R² values are 0.690 for “SWP” and 0.518 for “Ecological Integrity”. In PLS-PM, R² statistical significance is dependent on all incoming arrows (from the inner and outer models) and depends on the number of samples [9], which is forty here. For “SWP” there are 6 incoming arrows, while for “Ecological Integrity” there are 4. According to Hair Jr. et al., (2016) for a statistical significance of 5% the R² value of SWP must range between 0.5 and 0.75 for sample sizes between 39 and 48. In the model the figure is 0.618, which is within that interval. For “Ecological Integrity”, the R² value according to Hair Jr., et al., (2016) should also range from 0.5 to 0.75 since these values are for a sample size between 33 and 42. So it can be said that for the model represented in
Figure 3 both R² figures are significant. The following equations (1-6) show the measured score (MS) and predicted score (PS) for each latent variable in the “Overall” model.

\[
MS_{\text{Anthropogenic}} = \text{Population density} \times 0.059 + \text{cc}_e (ART)_{\text{with}} (WAB) \times 0.360 + \text{ci}_{pp} (ART) \times 0.606 + \text{ed} (ART) \times 0.082
\]  

(1)

\[
MS_{\text{Nature}} = \text{ed}_e (FOR) \times 0.658 + \text{e} (WAB)_{\text{with}} (FOR) \times 0.530 + \text{i} (FOR) \times 0.403
\]  

(2)

\[
MS_{\text{SWP}} = \text{BOD5} \times 0.343 + \text{Chlorides} \times 0.278 + \text{Conductivity} \times 0.305 + \text{Dissolved iron} \times 0.305
\]  

(3)

\[
MS_{\text{Ecological Integrity}} = \text{METI} \times 1.000
\]  

(4)

\[
PS_{\text{SWP}} = MS_{\text{Nature}} \times (-0.523) + MS_{\text{Anthropogenic}} \times (0.471)
\]  

(5)

\[
PS_{\text{Ecological Integrity}} = MS_{\text{SWP}} \times (-0.265) + MS_{\text{Nature}} \times (0.226) + MS_{\text{Anthropogenic}} \times (-0.351)
\]  

(6)

The “SWP” and “Anthropogenic” variables contribute negatively to “Ecological Integrity” because the corresponding path coefficients are negative at β = -0.265 and β = -0.351, respectively. These coefficients are negative because the increase in human pressure and the contamination of stream water cause losses of ecological integrity. The path coefficient linking “SWP” to Anthropogenic is positive (β = 0.471) as expected, since increasing anthropogenic pressures will increase stream water contamination. The variable “Nature” thus contributes negatively to “SWP” because the corresponding path coefficient is negative (β = -0.523). This makes sense, because if the state of the variables that make up "Nature" improves water quality then the concentration of physicochemical parameters in water will decrease. This highlights
the importance of riverside vegetation, which acts as a fundamental tool in good basin management, producing favourable impacts on water quality and serving as a buffer against contaminants (Jorgensen et al., 2000). “SWP” is formed by the physicochemical parameters, and those contributing negatively to “Ecological Integrity” are mainly conductivity ($W = 0.343$), chlorides ($W = 0.278$), dissolved iron ($W = 0.305$) and BOD$_5$ ($W = 0.305$).

![Figure 4.](image)

**Figure 4.** R-squared values of the PLS-PM model resulting from the average of SWP and METI for the respective hydrological years.

For each measured variable, the attributed weight seems to increase the latent variable value. All the weights are positive, but some variables have a stronger impact. In “Anthropogenic”, “ci_pp_(ART)” is the measured variable with most influence because it has the highest weight ($w=0.606$), while “cce_(ART)_with_(WAB)” has less impact ($w=0.360$). The percentage of artificial land uses beside water bodies ($W = 0.360$) and the percentage of connection of artificial land uses ($W = 0.606$) are the most heavily weighted measured variables in the latent variable “Anthropogenic”. The increase in edge density is a consequence of human intervention in the landscape (Uuemaa et al., 2007), which is why its degradation causes loss and fragmentation of the habitat and the deterioration of the biodiversity of ecosystems. Santos and Telleira (Santos and Tellería, 2006) relate the edge effect to the deterioration of habitat quality in regression, affecting the survival of the populations stationed in fragments. On the other hand, land use conflicts contribute to loss of soil fertility (Wu and Tiessen, 2002), contamination of surface and
groundwater (Ren et al., 2003) and loss of biodiversity (Martínez et al., 2009). The other two measured variables have a positive but near-zero influence with weights 0.059 for “Population density” and 0.082 for “ed_(ART)”. Population density and the edge density of artificial surfaces show a modest contribution of W = 0.059 and W = 0.082. Fernandes et al., (2019) analyses the role of population density and finds a greater contribution (W = 0.309), which they relate to the ineffective treatment of wastewater (Ferreira et al., 2017). The hydrological district of Galicia-Costa has a highly scattered population with a density of just 159.4 inhabitants/km². The low contribution of population density may be justified by the high level of dispersion in Galicia-Costa. On the other hand, this low influence could be due to the fact that, in the short term, the impacts of urban presence can be harder to show than on a longer time scale (Fernandes et al., 2019a). For “Nature”, the “ed_(FOR)” is the most substantial variable (W = 0.659) while the other two variables have a score that is still significant: “el_(WAB)_with_(FOR)” (w=0.530) and “lpi_(FOR)” (W = 0.403). In “SWP”, the variables seem to have an almost equivalent effect since all weights are near 0.3. “BOD₅” has the highest weight (W = 0.343), “Chlorides” has the lowest (W = 0.278) and “Conductivity” and “Dissolved iron” both have weights of 0.305. “Ecological Integrity” comprises a single measured variable, so by default the weight of METI is 1.000.

Aside from the direct effects (path coefficients) there are two more types of effect: indirect and total. Indirect effects are only established for indirect connections between latent variables. For example, “Nature” and “Anthropogenic” are directly connected to “Ecological Integrity”, but since both are connected to “SWP”, which is also connected to “Ecological integrity”. There is an indirect effect on “Ecological Integrity” from those variables. The indirect effect of “Nature” on “Ecological Integrity” is the product of the path between “Nature”-“SWP”(β=-0.523) and “SWP”-“Ecological Integrity”(β =-0.265). In this case the indirect effect is 0.139. Likewise, the indirect effect of “Anthropogenic” on “Ecological Integrity” is the product of (0.471) x (-0.265), -0.125. Since “SWP” is only connected to “Ecological Integrity” directly, the resulting indirect effect is zero. Total effects are simply the sum of direct effects and indirect effects. The total
effects of “Nature” and “Anthropogenic” on “SWP” are equal to the respective direct effects, i.e. -0.523 and 0.471, respectively. For the same reason, the total effect of “SWP” on “Ecological Integrity” is -0.265. The total effect of “Nature” on “Ecological Integrity” is 0.226+0.139=0.365 and that of “Anthropogenic” on “Ecological integrity” is -0.351-0.125=-0.476.

The present study is hard to compare to others because it covers a large number of basins with a fewer sampling points (Fernandes et al., 2018a)(Fernandes et al., 2019a). When several hydrographic basins are analysed, a lower R² may possibly be obtained than in studies that only contain sampling points in the same basin. When samples come from the same basin, the cause-effect relationships expressed may be more consistent, than when samples are scattered across different basins (Carter et al., 2019; Sanches Fernandes et al., 2018), which might result in a lower R2. Compared to another PLS-PM model (Fernandes et al., 2019a) optimised for prediction purposes, the overall statistical significance of weights and path coefficients is lower, so this model is not so suitable for prediction purposes. Despite this, this research presented here has the advantage of combining these important variables with other parameters relevant to water quality, having used a large number of sampling and measurement points. The combination of related anthropogenic pressures, water quality and ecological integrity in a large district such as the Galicia-Costa River Basin District, with 40 sampling points and over a long period of time, makes this analysis one of the most comprehensive conducted with PLS-PM in this area.

### 3.2. Partial Least Squares-Path Modelling

One of the main objectives of this manuscript is to examine the influence of the parameters used in PLS-PM and learn whether the cause-effect relationships can change from one hydrological year to another. Figure 4 shows the resulting R² values, which are almost constant across the models. For the models that represent the hydrological years of 2009-2010 the R² for “Ecological Integrity” is 0.399 and for 2014-2015 it is 0.470, which is significant only at a
significance level of 10%. The values for $R^2$ remain constant across the models, with the variable "SWP" being greater for all hydrological years than "Ecological Integrity" (Figure 4). The concept of ecological integrity has been used to manage aquatic systems, but it is considered difficult to quantify for managing terrestrial systems, particularly in wide areas (Carter et al., 2019). In our case, "Ecological Integrity" is affected by "SWP", so it makes sense that the $R^2$ for ecological integrity is lower. SWP represents the parameters on the water surface and negatively affects the quality of "Ecological integrity", except in 2014-2015, when it has a positive impact ($\beta = 0.096$). The resulting weights of variables are shown in Table 2. The total effects of variables on "SWP" remain largely constant over time, though small variations appear in each model. For "Ecological Integrity", "SWP" shows significant changes over time, and in 2014-2015 the effect actually becomes positive, which invalidates the respective weights of measured variables (Table 2) for that model since they are all positive. To perform a complete analysis of each measured variable, the weight was multiplied by the total effect on "Ecological Integrity" (Table 3), in order to learn the overall influence of each measured variable along with the model on "Ecological Integrity". The results shown in Table 3 reveal that 8 values do not match theoretical expectations (highlighted in yellow). These are the values of "Population density", which increases "Ecological Integrity" in 2009-2010 and 2010-2011, and "cce_(ART)_with_(WAB)" for 2013-2014 and 2014-2015. The effect of contaminant concentrations in hydrological year 2014-2015 is also unexpected, since all contaminants have a positive impact. This positive impact of the parameters measured in "SWP" is observed and positively affects "Ecology integrity" (Table 3), coinciding with the entry into force of Decree 1/2015 (Xunta de Galicia, 2015), which regulates the procedures for the preparation, approval and review of the Galicia-Costa hydrological plan, and deals with the assessment of damage to the public hydraulic domain or minor maintenance and conservation actions. A overall analysis of the changes in hydrological year 2014-2015 shows changes in "ed_(ART)" (Table 2). The edge density of artificial surfaces negatively affects the "Anthropogenic" variable for the hydrological years 2013-2014 and 2014-
2015, which is why it positively affects "Ecological Integrity". In other words, there is a trend towards improvement in the latest years of the study (Xunta de Galicia, 2015).

**Figure 5.** Total effects on "SWP" and "Ecological integrity" caused by the variables "Anthropogenic", "Nature", and "SWP" for each hydrological year.
Table 2. Weight of the parameters used in the PLS-PM model obtained for each year and overall.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Population density</td>
<td>0.059</td>
<td>-0.047</td>
<td>-0.034</td>
<td>0.098</td>
<td>0.106</td>
<td>0.242</td>
<td>0.172</td>
</tr>
<tr>
<td>cce_(ART)<em>with</em>(WAB)</td>
<td>0.36</td>
<td>0.425</td>
<td>0.452</td>
<td>0.396</td>
<td>0.348</td>
<td>0.131</td>
<td>0.276</td>
</tr>
<tr>
<td>ci_pp_(ART)</td>
<td>0.606</td>
<td>0.479</td>
<td>0.487</td>
<td>0.562</td>
<td>0.511</td>
<td>0.812</td>
<td>0.701</td>
</tr>
<tr>
<td>ed_(ART)</td>
<td>0.082</td>
<td>0.248</td>
<td>0.197</td>
<td>0.064</td>
<td>0.18</td>
<td>-0.103</td>
<td>-0.046</td>
</tr>
<tr>
<td>ed_(FOR)</td>
<td>0.658</td>
<td>0.712</td>
<td>0.682</td>
<td>0.696</td>
<td>0.645</td>
<td>0.738</td>
<td>0.528</td>
</tr>
<tr>
<td>el_(WAB)<em>with</em>(FOR)</td>
<td>0.53</td>
<td>0.535</td>
<td>0.518</td>
<td>0.493</td>
<td>0.488</td>
<td>0.524</td>
<td>0.54</td>
</tr>
<tr>
<td>lpi_(FOR)</td>
<td>0.403</td>
<td>0.305</td>
<td>0.374</td>
<td>0.374</td>
<td>0.457</td>
<td>0.265</td>
<td>0.571</td>
</tr>
<tr>
<td>BOD5</td>
<td>0.343</td>
<td>0.327</td>
<td>0.192</td>
<td>0.24</td>
<td>0.184</td>
<td>0.049</td>
<td>0.114</td>
</tr>
<tr>
<td>Chlorides</td>
<td>0.278</td>
<td>0.339</td>
<td>0.376</td>
<td>0.547</td>
<td>0.185</td>
<td>0.5</td>
<td>0.034</td>
</tr>
<tr>
<td>Conductivity</td>
<td>0.305</td>
<td>0.519</td>
<td>0.435</td>
<td>0.329</td>
<td>0.529</td>
<td>0.234</td>
<td>0.592</td>
</tr>
<tr>
<td>Dissolved iron</td>
<td>0.305</td>
<td>0.063</td>
<td>0.414</td>
<td>0.191</td>
<td>0.284</td>
<td>0.516</td>
<td>0.462</td>
</tr>
</tbody>
</table>
To test the statistical significance of PLS-PM models as regards effects and weights, a bootstrapping algorithm was used. The results are shown in Tables 4, 5 and 6 for weights, direct effects and total effects respectively. Of all the measured variables only “ed_(FOR)” and “el_(WAB)_with_(FOR)” remain statistically significant for all years. “Chlorides”, “Dissolved iron” and “Conductivity” are only statistically significant in certain periods. Even so, some of the total and direct effects could reach statistical significance (Tables 5 and 6).

The effect of “Nature” and “Anthropogenic” on “SWP” is undoubtedly consistent since it is found to be statistically significant for all models in all years. For “Ecological Integrity”, “Anthropogenic” remains a significant variable, but it is not significant for the hydrological years between 2009 and 2011. “Nature” is only statistically significant for hydrological year 2014-
2015, while “SWP” is had significant for 2010-2011 and 2012-2013. Table 3 shows the statistical significance of the total effects of latent variables on “Ecological Integrity”.

**Table 4.** Statistical significance of weights, with all values with a probability higher than 0.05 marked in green, to reject the null hypothesis.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Population density</td>
<td>0.796</td>
<td>0.894</td>
<td>0.903</td>
<td>0.677</td>
<td>0.66</td>
<td>0.337</td>
<td>0.527</td>
</tr>
<tr>
<td>cce_(ART)<em>with</em>(WAB)</td>
<td>0.16</td>
<td>0.116</td>
<td>0.091</td>
<td>0.107</td>
<td>0.231</td>
<td>0.692</td>
<td>0.329</td>
</tr>
<tr>
<td>ci_pp_(ART)</td>
<td>0.078</td>
<td>0.215</td>
<td>0.163</td>
<td>0.085</td>
<td>0.14</td>
<td>0.067</td>
<td>0.061</td>
</tr>
<tr>
<td>ed_(ART)</td>
<td>0.762</td>
<td>0.472</td>
<td>0.512</td>
<td>0.8</td>
<td>0.504</td>
<td>0.738</td>
<td>0.857</td>
</tr>
<tr>
<td>ed_(FOR)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.001</td>
</tr>
<tr>
<td>el_(WAB)<em>with</em>(FOR)</td>
<td>0</td>
<td>0.002</td>
<td>0</td>
<td>0.001</td>
<td>0.001</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>lpi_(FOR)</td>
<td>0.012</td>
<td>0.189</td>
<td>0.059</td>
<td>0.026</td>
<td>0.013</td>
<td>0.105</td>
<td>0.002</td>
</tr>
<tr>
<td>BOD5</td>
<td>0.088</td>
<td>0.229</td>
<td>0.159</td>
<td>0.146</td>
<td>0.401</td>
<td>0.842</td>
<td>0.687</td>
</tr>
<tr>
<td>Chlorides</td>
<td>0.276</td>
<td>0.23</td>
<td>0.095</td>
<td>0.005</td>
<td>0.493</td>
<td>0.059</td>
<td>0.93</td>
</tr>
<tr>
<td>Conductivity</td>
<td>0.25</td>
<td>0.063</td>
<td>0.058</td>
<td>0.112</td>
<td>0.03</td>
<td>0.289</td>
<td>0.032</td>
</tr>
<tr>
<td>Dissolved iron</td>
<td>0.045</td>
<td>0.777</td>
<td>0.002</td>
<td>0.187</td>
<td>0.219</td>
<td>0.013</td>
<td>0.046</td>
</tr>
</tbody>
</table>

**Table 5.** Statistical significance of direct effects, with all values with a probability higher than 0.05 marked in green, to reject the null hypothesis.
As shown in Table 4, “Anthropogenic” and “Nature” are statistically significant for all models, which is not the case for “SWP”. This shows that the effects of anthropogenic pressures and of forestry environments have a constant significance over time on ecological integrity, while the forms and impact of contamination can vary from one hydrological year to another.

Table 6. Statistical significance of total effects on “Ecological Integrity”, with all values with a probability higher than 0.05 marked in green, to reject the null hypothesis.
3.3. General discussion

The population density used is the average population density of each drainage area. In this case, with the sampling points so widely dispersed throughout the study area and with the widely scattered population, this may result in a positive effect on “Ecological Integrity”. Hydrological years 2009-2010 and 2010-2011 are cases in point. To improve the model, more sampling points would need to be added for each basin, especially points near population centres. This would mitigate the effect of population dispersion.

The models studied here present the relationships that affect the ecological integrity of the basins that make up the study area as well as the surface quality of the water. Applying these models could serve as an aid for the administration in making the right decisions. In this way, those variables with most weight can be detected, and those for which corrective measures need to be carried out. These models can therefore act as a tool for guaranteeing information for optimal governance of water resources as well as for facilitating the integrated management of basins via the analysis of all the parameters necessary to address their political, social, and economic systems. With the proper use of this tool, water security could be guaranteed and sustainable development could be promoted. Encouraging the use of these models would supply information conducive to the correct management of the basins, encouraging their coordinated, joint organisation. This integrated watershed management in turn is conducive to the sustainable development of these systems.

A correlation between the decline in ecological integrity and in increase in human population density across rural–urban landscape gradients has been reported in various studies (Quigley et al., 2001; Urban et al., 2006). These studies highlight the need for ecological restoration measures as the main objective in the administrative management of the basins and sub-basins (Horak et al., 2019). It has also been detected that ecological integrity can be improved through reforestation (Perino et al., 2019). As analysed in this study, improvements in riparian vegetation
and changes in land use in regard to the forest environment would improve ecological integrity. Moreover, the quality of water, which decreases the ecological integrity of ecosystems, is affected by numerous variables (Stone et al., 2005). Multiple studies agree that changes in riparian vegetation (Dosskey et al., 2010; Souza et al., 2013) and high population densities with urban, industrial and intensive agriculture settlements cause great pressure on surface water bodies and consequently lead to deterioration in water quality.

It is possible to improve these models, for example by refining the monitoring network, because in general they only have an average of two sampling points per river. Increasing the number of points would make the monitoring network more representative. Moreover, with more points, sources of contamination could be accurately detected, as could incorrect measurements due to instrument failure or human error. In addition, the real state of riparian vegetation needs to be assessed, as it is a fundamental variable that influences the quality of aquatic ecosystems: the degradation of this protective barrier leads to a loss of ecological integrity and consequently to a decrease in biodiversity (Urban et al., 2006). Also, the damage caused by human pressure often has serious consequences for human health and for the local economy (Horak et al., 2019).

Population is another relevant parameter, which in this case affects the “Anthropogenic” latent variable. In this case the population density does not contribute the expected weight due to how scattered the population of the Demarcation is. An alternative would be to analyse how the points with higher population densities affect the model. In this way, these areas would better reflect the real state of the basins to which they belong. In turn, adding point source emission information would provide the model with significant references on how they really affect the ecological integrity of the basins and on what parameters positively affect this variable.

Finally, it is worth noting that this predictive model could be substantially improved by including climate variables. This could help to provide a tool that would enable changes and adaptation improvements to be made as regards climate change. The use of partial least squares in path
models results in a promising tool for assessing the state of the water bodies of the entire Galicia-Costa Demarcation.

4. Conclusions

Partial least squares -path modelling (PLS-PM) is applied in the Galicia-Costa hydrological district, consisting of 18 basins. Different pressures and consequences for ecological integrity are studied. The application of these models provides an overview of the factors that may initially be considered as important for optimal water governance. Of the latent variables, "Anthropogenic" is strongly influenced by the percentage of artificial land use, which is highly notable for its influence on surface water parameters (SWP). The PLS-PM results show that the loss of ecological integrity in these basins is largely due to the “Anthropogenic” latent variable. In the analysis of the model over the years, an improvement in the parameters measured by the latent variable SWP stands out. In turn, an improvement is detected in the "Edge density of artificial surfaces". All this is beneficial for "Ecological Integrity". In view of the results obtained here, it is considered relevant to improve the riverside vegetation, since this would indirectly improve water quality in terms of both physicochemical and biological parameters, thereby facilitating the integral management of the basins. On the other hand, land use planning policies must work in conjunction with technological solutions to improve water quality.

In terms of the management of hydrographic basins, this study analyses a large number of sampling points, identifies the pressures that degrade ecological integrity and thus presents an effective modelling tool for developing the water governance policies needed to guarantee the water security of these water bodies. Overall, the PLS-PM model finds strong links between the different variables and constitutes a promising tool for facilitating protection and correction measures and thus facilitating decision-making that can guarantee water security. The next step
should be to make small improvements in the model, such as increasing the number of sampling points or information on discharges.

5. Acknowledgments

This research was funded by the Consellería de Educación, Universidade e Formación Profesional, Xunta de Galicia, España, under project R815 131H 64502. The authors thank Augas de Galicia, belonging to the Xunta de Galicia, for their collaboration.

The authors belonging to the CITAB Research Centre were also funded by the ERDF/COMPETE/POCI—Operational Competitiveness and Internationalisation Programme, under Project POCI-01-0145-FEDER-006958, and by the National Funds of FCT—Portuguese Foundation for Science and Technology, under the project UID/AGR/04033/2020. For the author belonging to the CQVR, the National Funds of FCT—Portuguese Foundation for Science and Technology, under the project UIDB/QUI/00616/2020 and UIDP/00616/2020 also supported the research. Financial support was also provided by the FCT-Portuguese Foundation for Science and Technology (Grant: SFRH/BD/146151/2019) to António Fernandes.

Legend of Tables:

Table 1. List of measured variables used as source data for Partial Least Squares–Path Modeling (PLS-PM). Besides identifying and describing variables and giving the units in which they are measured, the table gives indications on usage in the PLS-PM models and on the data sources.

Table 2. Weight of the parameters used in the PLS-PM model obtained for each year and overall.

Table 3. Product of the weights of each parameter for the total effect on ecological integrity.
Table 4. Statistical significance of weights, with all values with a probability higher than 0.05 marked in green, to reject the null hypothesis.

Table 5. Statistical significance of direct effects, with all values with a probability higher than 0.05 marked in green, to reject the null hypothesis.

Table 6. Statistical significance of total effects on “Ecological Integrity”, with all values with a probability higher than 0.05 marked in green, to reject the null hypothesis.

Legend of Figures:

Figure 1. Study area: watersheds analyzed

Figure 2. Land use and population density in study area

Figure 3. Diagram of complete PLS-PM models for Galicia-Costa. The circles represent the latent variables, namely “Anthropogenic”, “Nature”, “SPW” and “Ecological integrity”. The rectangles represent formation variables (see description in Table 1). The arrows represent the links between formation and associated latent variables, and between related latent variables, while arrow labels are weights and path coefficients that quantify those links.

Figure 4. R-squared values of the PLS-PM model resulting from the average of SWP and METI for the respective hydrological years.

Figure 5. Total effects on ”SWP” and ”Ecological integrity” caused by the variables ”Anthropogenic”, ”Nature”, and ”SWP” for each hydrological year

6. References


European Commission, 2000. Directiva 2000/60/CE del Parlamento Europeo y del Consejo, de 22 de diciembre de 2000, por la que se establece un marco comunitario de actuación en el ámbito de la política de aguas. Directiva 2000/60/CE.


Galicia, X. de, 2015. DECRETO 1/2015, de 15 de enero, por el que se aprueba el Reglamento de la planificación en materia de aguas de Galicia y se regulan determinadas cuestiones en desarrollo de la Ley 9/2010, de 4 de noviembre, de aguas de Galicia. Cons. medio Ambient. Territ. e infraestructuras 10, 2256.


IGE, 2015. Instituto Galego de Estatística, “Datos poboacionais por concellos”.


