To cite this article: Requejo-Castro, D., Giné-Garriga, R., Pérez-Foguet, A., 2019. Bayesian network modelling of hierarchical composite indicators. Sci. Total Environ. 668, 936–946.

To link to this article: https://doi.org/10.1016/j.scitotenv.2019.02.282

To download this article (free access until May 01, 2019): https://authors.elsevier.com/a/1YibgB8cckLuU

BAYESIAN NETWORK MODELLING OF HIERARCHICAL COMPOSITE INDICATORS

David Requejo-Castro[* 1], Ricard Giné-Garriga[** 2], Agustí Pérez-Foguet[* 3]

[*] Engineering Sciences and Global Development (EScGD), Department of Civil and Environmental Engineering, Universitat Politècnica de Catalunya · BarcelonaTech (UPC), Jordi Girona, 1-3, 08034 Barcelona, Spain

[**] Stockholm International Water Institute (SIWI), Linnégatan 87A, 100 55 Stockholm, Sweden

Email[1]: david.requejo@upc.edu

Email[2]: ricard.gine@siwi.org

Email[3]: agusti.perez@upc.edu

Corresponding Author: David Requejo-Castro

Address: Universitat Politècnica de Catalunya - Barcelona School of Civil Engineering

CAMPUS NORD - Edif. C2

C. JORDI GIRONA, 1-3

08034 BARCELONA

SPAIN

Email: david.requejo@upc.edu

Telephone: (+34) 665 14 18 29

Abstract

The water, sanitation and hygiene (WaSH) sector has witnessed the development of multiple tools for multidimensional monitoring. Hierarchical and composite indicators (CI)-based conceptual frameworks provide one illustrative example. However, this approach does not address the existing interrelationship of the indicators they integrate. Bayesian Networks (BNs) are increasingly exploited to assess WaSH issues and to support planning and decision-making processes. This research aims to evaluate the validity, reliability and feasibility of BNs to replicate an existing CI-based conceptual framework. We adopt a data-driven approach and we propose a semi-automatic methodology. One regional monitoring initiative is selected as a pilot study: the Rural Water Supply and Sanitation Information System (SIASAR). Data from two different countries are processed and analysed to calibrate and validate the model and the method. Major findings show i) an improvement of model inference capacity when providing structure to the networks (according to the CI-based framework), ii) a reduction and quantification of the key components that explain a pre-defined objective variable (implying important advantages in data updating), and iii) an identification of interlinkages among these components (which might enhance multi- and trans-disciplinary actions). We conclude that BNs accurately replicates the CI-based conceptual framework. The proposal contributes to its wider application.

Keywords

Bayesian Networks; composite indicators; systematic methodology; interlinkages; water, sanitation and hygiene

1. INTRODUCTION

Indicators have played a key role in monitoring and evaluation, reporting, decision- and policy-making, and public communication. An important feature of indicators is their capacity to summarise, focus and condense information about complex systems integrating environmental, social and economic aspects (Godfrey and Tood, 2001). However, a solely indicator cannot capture the complexity of the real world and, much efforts has thus gone into the development of a wide range of related approaches. For instance, composite indicators (CI) have been widely used to evaluate the existing multi-dimensionality of the problems at hand (Nardo et al., 2005). Some of these CI have been structured hierarchically (i.e. indicators, sub-indices and indices) based on conceptual frameworks defined by experts (several examples in Bandura, 2006; Singh et al. 2009) and enhancing multi- and cross-disciplinary approaches. In particular, the water, sanitation and hygiene (WaSH) sector has witnessed the development of multiple alternatives towards the multidimensional monitoring of WaSH issues. The WHO and UNICEF Joint Monitoring Programme for Water Supply, Sanitation and Hygiene (JMP) has been taken over the role at international level of reporting on the status of access to water and sanitation, shifting from technology-based indicators to multidimensional measures of the level of service delivered (Joint Monitoring Programme, 2000; 2008; 2017). From a more academic perspective, the use of CI appeared as a helpful tool to evaluate WaSH aspects from many disciplinary perspectives and conceptual frameworks. Thus, it is possible to find composites which tackle independently water-related (Bordalo et al., 2006; Cho et al., 2010; Cohen and Sullivan, 2010; Giné-Garriga and Pérez-Foguet, 2010, 2011; Sullivan et al., 2003), sanitation-related (Giné-Garriga et al., 2017; Giné-Garriga and Pérez-Foguet, 2018; WSP, 2015) or hygiene-related issues (Giné-Garriga and Pérez-Foguet, 2013; Webb et al., 2006). Additionally, more integrated approaches have addressed WaSH-related issues from a Human Right perspective (Flores Baquero et al., 2016; Luh, et al., 2013) and from a more sectoral-focused one (Giné-Garriga and Pérez-Foguet, 2013; Godfrey et al., 2014; Requejo-Castro et al., 2017). One major weakness is that CI do not address the existing interrelationship of the indicators they integrate; they do not describe the increasing interdependency of the real world.

This drawback has been widely tackled by applying data-driven and conceptual approaches. Within the former, classical techniques such as Principal Component Analysis (PCA) and Factor Analysis (FA) have been used to highlight the statistical relationships between indicators in terms of a smaller number of components or factors, respectively (Nardo et al., 2005). These techniques are based on the variance of a given data set. While PCA reduces the dimensionality of a data set by finding correlated variables, FA determines the number of latent independent variables underlying the data. On the other hand, the principle of causality has been the main feature introduced in the so-called "conceptual" approach, which have been mainly applied in the environmental field. They are based on two different approximations, namely causal chains and causal networks. Within the former, most of the publications apply the Pressure-State-Response framework (PSR; OCDE, 1993) and its transformations: Driving force–State–Response (DSR; UN, 1996) and Driving force–Pressure–State–Impact–Response (DPSIR; EEA 1999). Here, the main idea is to identify, for instance, the progressive chain of events, from driving force and related pressure indicators, leading to state change, impact and response ones. On the other hand, other authors have suggested the appropriateness of causal networks within the DPSIR framework rather than causal unidirectional chains to deal with the complexity of real world connections (Lin et al., 2009; Niemeijer and De Groot, 2008). In addition to this, few studies have applied techniques such as the Analytic Network Process (Saaty, 2001) or PCA and FA to quantify, in terms of weight, indicators' linkages (Hou et al., 2014; Vacik et al., 2007; Wolfslehner and Vacik, 2011). Considering both causal chain (mostly DPSIR framework and its derivatives) and causal networks uses, these approaches have been helpful to conceptual framework development (Chandrakumar and McLaren, 2018; Ortiz-Lozano, 2012; Taft and Evers, 2016; Wolfslehner and Vacik, 2011), to scenario assessment (Chung and Lee, 2009; Ramos-Quintana et al., 2018; Scharin et al., 2016) and to structure modelling exercises (García-Santos et al., 2018; Pirrone et al., 2005). The PSR conceptual framework has been also applied to address the linkages between water scarcity and poverty (Pérez-Foguet and Giné-Garriga, 2011).

Bayesian Networks (BNs) have been exponentially used (Aguilera et al., 2011; Marcot, 2017) to explore the interdependencies and cause-effect relationships, simulating complex problems that involve a large number of variables that are highly interlinked. Thus, BNs raise as a complementary approach to tackle above mentioned drawbacks. Briefly, BNs are probabilistic graphical models where the conditional dependencies of the variables relevant to a particular study are encoded within directed acyclic graphs (DAGs). Each node of the graph is associated with one variable of the data set. The directed links connecting the nodes represent informational or cause-effect relationships. These dependencies are quantified by the conditional probability tables (CPTs), which represent the extent to which one node is likely to be affected by the others.

Apart from facilitating common features to the first two approaches presented (i.e. CI and causal chains and networks) such as scenario analysis, BNs have been interestingly used to identify the key factors influencing aspects of interest (Li et al., 2019; Song et al., 2018) and to elucidate the network structure underlying the data at hand through associated structure learning algorithms (SLA) (Alameddine et al., 2011; Garcia-Prats et al, 2018). BNs have been successfully applied to address environmental issues (Bromley, 2005) and water issues (Phan et al., 2016), but their application to the WASH sector is less common (Cronk and Bartram, 2017, 2018; Dondeynaz et al., 2013; Fisher et al., 2015; Giné-Garriga et al., 2018; Kumar and Mazumdar, 2002). By way of examples, Cronk and Bartram (2018) applied BNs to define the factors influencing 24 hours water service availability and exploring which of the variables considered were more influential in water discontinuity; and Giné-Garriga et al. (2018) developed a BN system to monitor WaSH national programmes, taking as a reference point a WaSH-related multidimensional index (Giné-Garriga and Pérez-Foguet, 2013).

This paper first seeks to expand and deepen the knowledge on the validity, reliability and accuracy of BNs to replicate CI-based conceptual frameworks. Secondly, it provides a systematic methodology for network construction to replicate the structure of composite indicators that are organized hierarchically. We adopt a data-driven approach. One regional monitoring initiative is selected as a pilot study: the Rural Water Supply and Sanitation Information System (SIASAR). The SIASAR's conceptual framework comprises a set of composite indicators (16), which are aggregated into different dimensions of interest (4), partial indices (2) and a general index. Specifically, we process and analyse data from two different countries, namely Nicaragua and Honduras.

This paper is structured as follows: Section 2 describes the selected pilot study in detail. A step-by-step explanation of the proposed methodology to deal with networks generation is conducted in Section 3. Main results regarding the method are presented and discussed in Section 4, including some application examples. Major findings are highlighted in Section 5 to conclude the study.

2. SELECTION OF THE PILOT STUDY

This study exploits the conceptual framework of SIASAR, one information system widely implemented in Latin America and Caribbean (LAC). We opted for this pilot study due to i) its open-database available for interested stakeholders, and ii) its complex hierarchical structure of indicators and composite indicators.

SIASAR collects primary data and after a systematic process of data validation, datasets are published and can be easily accessed online (Requejo-Castro et al., 2017; SIASAR, 2017). For data collection, a set of questionnaires are developed to analyse the sustainability of services from four different perspectives: i) the community, ii) the water system, iii) the service provider, and iv) the technical assistance provider. Primary data are then validated prior their publication through i) systematic data checks, and ii) ad-hoc validation by the municipal or national authority (SIASAR, 2018).

As regards the second aspect, SIASAR's conceptual model is made up of six aggregated dimensions and three additional indices. Nevertheless, by the time of this research, data was only available for four of the six dimensions. Thus, we consider a simplification of the conceptual model (see Table 1).

Table 1. General index, partial indices, dimensions and components of the SIASAR simplified conceptual model. Source: Requejo-Castro et al., 2017.

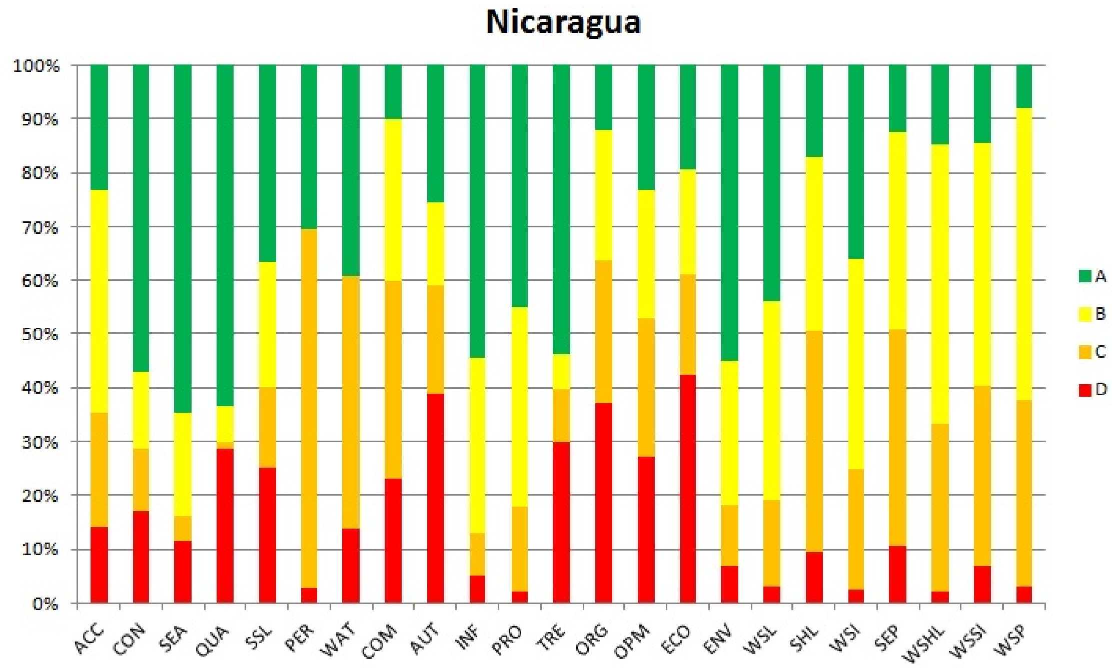
Water and Sanitation Service Performance Index (WSP)	
Water, Sanitation and Hygiene Service Level (WSHL)	**Water Services Sustainability Index (WSSI)**
Water Service Level (WSL)	**Water System Infrastructure (WSI)**
Accessibility (ACC)	System Autonomy (AUT)
Continuity (CON)	Production Infrastructure (INF)
Seasonality (SEA)	Water Catchment Protection (PRO)
Quality (QUA)	Treatment system (TRE)
Sanitation and Hygiene Service Level (SHL)	**Service Provision (SEP)**
Sanitation Service Level (SSL)	Organizational Management (ORG)
Personal Hygiene (PER)	Operation & Maintenance Management (OPM)
Household Hygiene (WAT)	Economic Management (ECO)
Community Hygiene (COM)	Environmental Management (ENV)

Each dimension comprises four components, which are fed by a short list of single indicators (see Appendix Table A1). As collected data are frequently represented on different scales, these indicators are normalized prior to their analyses. A score between 0 and 1 is assigned for each parameter, whereby 1 represents the best performance and 0, the worst-case scenario. Components are then defined by simple and easy-to-use multi-attribute utility functions (Pérez-Foguet and Flores-Baquero, 2015). In Table 2, we provide two illustrative examples. The first one addresses criteria whose values are associated with linear variations (i.e. water coverage). In this case, an additional indicator, such as the distance to the water source, is considered to evaluate the overall value of the component "Accessibility". The second example takes into account several criteria to assess the component "Operation & Maintenance (O&M) management" of the service provider. First, a different punctuation is given according to the residual chlorine measurement obtained. Second, a combination of different O&M aspects are assessed and rated. A final value of the utility function is provided by a linear mean of both criteria. Once the components are calculated, the different composites are constructed by aggregating the components arithmetically, in the case of the four dimensions, or geometrically for the partial and general indices (see Appendix Table A.2).

Table 2. Examples of utility functions associated to the components of Accessibility (up) and Operation & Maintenance management (bottom).

Accessibility (ACC)				
Effective coverage	Coverage = 0	Coverage * Accessibility factor Acc. fact. = 1 (if average dist. ≤ 100 m) Acc. fact. = 2/3 (if average dist. > 100 m)		Coverage = 1 Average dist. < 100 m
Criteria	**0**	**Linear variation**		**1**
		0.33	**0.66**	
Chlorine basic operation (residual chlorine)	Cl ≤ 0.1	0.1 mg/l < Cl ≤ 0.3 mg/l	Cl > 1 mg/l	0.3 < Cl ≤ 1. mg/l
General assessment of O&M	Neither corrective nor preventive maintenance is provided	Corrective AND/ OR preventive maintenance is provided	Corrective AND/ OR preventive maintenance is provided, AND O&M costs are registered	Corrective AND preventive maintenance is provided, AND O&M costs are registered, AND a plumber is present at the organization
Operation & Maintenance Management (OPM)				

Finally, the achieved results are made more understandable for final users and stakeholders by linking components, dimensions and index values to a defined grading system (A, B, C or D, whereby A represents the best result and D, the worst). Although, in practice different intervals are used, in this study equal intervals are employed. Specifically, these intervals are defined as follows: A, [1–0.75], with both limits included; B (0.75–0.5]; C (0.5–0.25]; and D (0.25–0]. This aspect is especially important as they represent the states of the variables (nodes) or, in other words, the way they are parameterised.



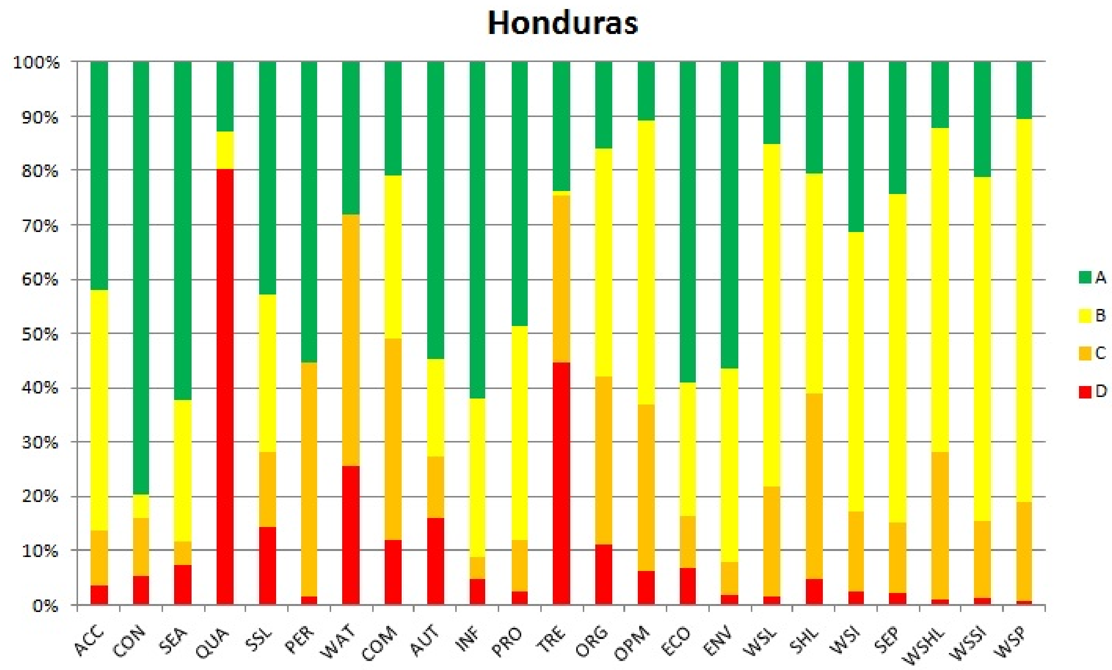


Figure 1. Components, dimensions, partial indices and general index "ABCD" distributions.

In this research, we selected in first instance the case of Nicaragua, where a baseline of all rural communities, water systems and service providers was carried out. Here, and after data pre-processing, a total number of 3,495 communities counting with values associated to all components detailed in the simplified conceptual model is employed. Additionally, the database of Honduras is exploited as a second case study, where 3,608 communities fulfil the same requirement (both data sets available at http://doi.org/10.5281/zenodo.804010).

In Figure 1, we show the distribution of the primary data according to the above mentioned intervals or states. In general, and for both cases, all components possess values associated with each state. Exceptions exist for "personal hygiene" (PER) and "household hygiene" (WAT), where values belong to A, C and D states due to the utility function structure. Specifically, in Nicaragua (see Appendix Table A.3), the average distribution of A, B, C and D states are 38.3, 23.0, 21.3 and 20.3, respectively. However, important differences are found when focusing on individual states. For instance, a 64.7% of the rural communities are evaluated as "A" considering the component of "seasonality" (SEA), while just a 10.1% reach this qualification as far as "community hygiene" (COM). Additionally, it is possible to identify those components in a more precarious situation (i.e. the ones related to hygiene and service provider performance). In Honduras, the average distribution is 42.2, 27.3, 18.7 and 15.2 for the states A, B, C and D, respectively. In this case, there are wider differences within the states, such as "D", where there is a maximum of 80.2 and a minimum of 1.5. Differently than Nicaragua, here those components related to "water quality" (QUA) and "treatment system" (TRE) require more attention. In brief, these two countries represent two different realities and specificities of the same sector of our interest, thus making it suitable for the purpose of this work.

3. METHODS

This Section proposes a step-by step methodology to replicate the SIASAR hierarchical CI-based framework by exploiting the flexibility of BNs. First, however, we present the main novelty of this work in terms of method, and we introduce the main assumptions of the approach adopted.

There are mainly two approaches to develop a BNs model: manual or automatic (Aguilera et al., 2013). The first one includes expert opinion and complementary literature as part of the process to define which variables are dependent to each other and to which extent. The second approach (i.e. automatic) involves the use of structure learning algorithms (SLA), which provides the optimum structure of the network spurred on by data. This data-driven approach can reduce the subjectivity of the decisions to make when defining the structure of the network and can also help to minimize the need for expert elicitation, which it is time consuming process (Alameddine et al., 2011). To the best of our knowledge, none of the previous studies in the WaSH sector has proposed a data-driven approach when applying BNs modelling. In model construction, three key assumptions have been made. First, this method is applied to large set of databases. Second, a limited number of states are used for each variable. Specifically, we employ the "A-B-C-D" grading system that characterized the presented pilot study. Third, we only allow variables connecting others which are at the same or the next hierarchical level.

With regard to method for BN construction, it is illustrated in Figure 2 and briefly outlined below:

3.1. Initial settings

First, it is important to highlight the existence of an important amount of both free and commercial software available. Furthermore, existing software employs one of the three types of SLA, namely constraint-based, score-based and hybrid approaches (Liu et al., 2017; Madsen et al., 2016). Here, we propose the use of R free software and its package "bnlearn (version 4.1)" developed by Scutari (2010). In this case, constraint-based and score-based methods for network structure learning are available. Briefly, the former learns the network structure by analyzing the probabilistic relations with conditional independence (CIn) tests and the latter assigns a score to each candidate BNs and try to maximize it with some heuristic search algorithm (Scutari, 2010). Further details of these algorithms are not provided in this study, but have been extensively described by Liu et al. (2017). As large amount of data is needed to guarantee the reliability of the CIn tests, the selected databases are suitable for applying the constraint-based algorithms.

3.2. Bayesian Networks model generation

Once the initial settings have been introduced, the proposed methodology for BNs model generation comprises four main stages (see Figure 2): model proposal (A), model calibration (B), model validation (C) and method validation (D). Each stage integrates different steps which are specified as follows:

- *Step 0. Conceptual model selection and data split*

The first step falls on selecting an appropriate conceptual framework to test the proposed methodology. We suggest those ones that possess, at least, three levels of variables (e.g. indicators, partial indices and general index). Then, we propose following other authors recommendation and randomly splitting the data into two parts, one for training the model and the other for testing. This process, which is directly related to model evaluation, helps ensure that the final result is feasible and defensible (Chen and Pollino, 2012). In this case, we opted for assigning the 70% of the dataset for training and the 30% for the testing. The first is used for the stages A and B of the methodology.

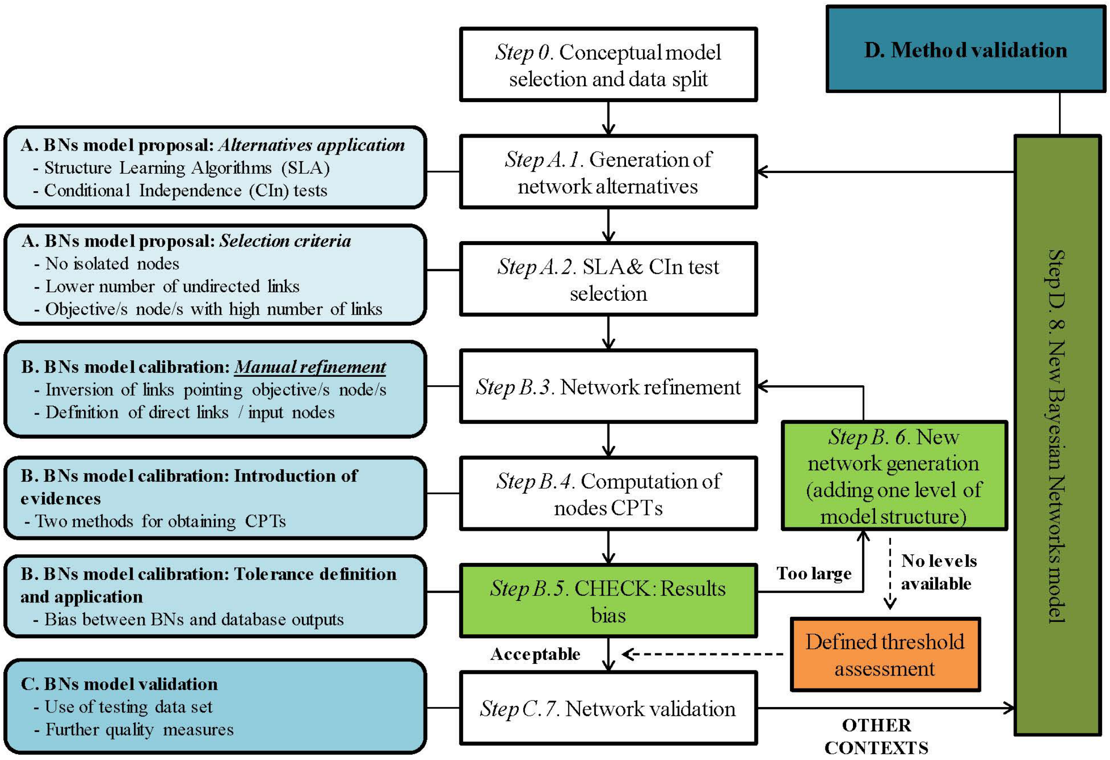


Figure 2. Step-by-step methodology for BNs model generation and validation.

- *Step A.1: Network alternatives generation*

"bnlearn (version 4.1)" package implements the following constraint-based learning algorithms (Scutari, 2010): grow-shrink (gs), incremental association (iamb), fast incremental association (fast.iamb), interleaved incremental association (inter.iamb) and max-min parent and children (mmpc). Additionally, the CIn tests must be chosen regarding data typology. In this case, all values are discrete and several CIn tests for discrete values are available. Regarding computing time, CIn tests selection is reduced to mutual information (mi) and Pearson's X^2 (x2) tests. By applying the combination of both SLA and CIn tests to the database, ten primary networks are obtained.

- *Step A.2: Structure Learning Algorithm (SLA) + Conditional Independence (CIn) test selection*

Alameddine et al. (2011), while using constraint-based algorithms to learn the structure of the network from the data, suggests comparing different potential networks by evaluating model's ability to incorporate the relevant endpoint (understood as the answer of interest) or to identify variable links. In this sense, and in order to select the final primary network (i.e. SLA and CIn test tandem), we propose several joint criteria. Thus, the final selection should be that one which i) does not leave isolated an important number of nodes, ii) presents a lower number of undirected links, and iii) identifies a higher number of links associated with the objective node(s). Regarding this last aspect, and as a means of example, node "WSP" (general index) is considered as the objective one, according to the hierarchical structure of the presented conceptual model.

- *Step B.3. Network refinement*

Once the primary network is selected, and in contrast to the previous process which can be carried out automatically, a further manual refinement is required. This process, as the first part of the BNs model calibration, comprises several actions:

i) Inversion of links associated with the objective node. It might occur that the link between the objective node and any other node goes from the former to the latter. Considering the structure of the conceptual framework at hand, this must be solved by reversing this direction;

ii) Definition of direct links. BNs can only be fully operative (e.g. able to simulate scenarios) in the absence of undirected links. Otherwise, a complete directed acyclic graph (DAG) couldn't be obtained. In this sense, we recommend the decision of providing this direction to be supported by expert consultation;

iii) Definition of input nodes. We propose three alternatives for the final selection of the network input nodes. First, a "data-driven approach" is put forward. In this case, inputs nodes are selected by providing a minimum manipulation to the network (input nodes proposed by SLA and CIn test). Second, an "interquartile range (IQR) approach" is proposed. Here, input nodes and intermediate ones (those between the input and objective nodes) are represented by those variables with highest IQR (see Appendix Table A). Third, a "smart variable approach" is taken into consideration, where input nodes are selected according to their facility for data collection in field. In any case, this manipulation must be done carefully and some decisions must be taken in order to avoid closed loops. At this point, four different functional networks are obtained for further analysis.

- *Step B.4. Computation of nodes CPTs*

Nodes CPTs represent the probability of each possible state in a "child" node given each possible event in the "parent" node(s). In this study, node states are represented by an "A-B-C-D" grading system (inherited from the selected conceptual model). Thus, every node is associated with a probability distribution according to these states. With the aim to simplify the nomenclature, the notation for model results will be denote as "inferred distributions" and the ones provided by the available data as "IS (information system) distributions".

Having said this, and once networks are functional, a "direct use" (forward direction from input nodes to objective node) is carried out. This application is known as predictive inference (Carriger et al., 2016). In doing so, evidences are introduced to compare inferred distributions with IS distributions elicited from data. Specifically, IS distributions (obtained from Nicaragua database) are assigned to the corresponding input nodes. Then, "WSP" node inferred distributions are compared with the IS distributions provided by the data. This procedure is carried out for the different networks at hand. We also suggest testing the two methods available to compute the CPTs of any node of interest, namely "maximum likelihood estimates (mle)" and "Bayesian setting (bayes)" (Scutari and Denis, 2015).

- *Step B.5. Results bias check*

The last step regarding the calibration process falls on the assessment of each network inferred results. A priori, a threshold bias between 0% and 5% is established. Here, two possible results might come up. First, one, two or all networks provide results within the threshold. Then, that network providing a lower bias (lower difference between inferred and IS distributions) should be selected. In those cases where results are similar, it is proposed to employ the minimum mean bias as to carry out the final network selection. Finally, the procedure should jump to Step C.7 (BNs model validation).

On the other hand, if resulted bias is larger than the threshold established, an iterative intermediate step should be carried before model validation as following detailed.

- *Step B.6. New network generation*

When the bias obtained is not acceptable, it is proposed to provide structure to the network. In this case, structure is related to the conceptual model selected which, in other words, represents the expert knowledge contribution during its development. As an illustrative example, previous steps could be carried out by setting a first scenario employing those sixteen components of the conceptual model and the general index. If the bias is too large, a second scenario should be simulated introducing the two partial indices as new nodes (WSHL and WSSI), complementarily to the sixteen components and general index. Finally, the bias should be checked again. In the case the bias is still large, the loop involving Steps 3-5 should be repeated (adding a new level of model structure) till an acceptable result is achieved. Additionally, if there is not the possibility to add a new level of structure to the network and the bias is still too large, it is recommended to assess whether the suitability of the threshold established and the subsequent trade-offs.

- *Step C.7. Network validation*

As to validate the selected network, different actions are suggested as well:

i) Check using testing data set. Similar to Step B.4, evidences are introduced to compare inferred distributions with the ones elicited from data. Instead, and based on the network obtained, the dataset kept for testing is used. Then, result biases are checked. Optionally, same procedure might be done by using the complete dataset without splitting.

ii) Assessment of further quality measures. Apart from evaluating the bias obtained as regards the objective node, we recommend to assess as well the inferred distributions associated to those nodes which provide structure to the model. Then, they should be compared to the associated real distributions.

- *Step D.8. Method validation*

Last but not least, a final step is included in this study with the aim to validate the proposed methodology. This step might not be possible when dealing with any other conceptual framework, as additional data from another context is required. In such a case, it is recommended to focus on the results obtained in previous steps. Here, it was possible to access to another country data to validate the methodology, which it is representative in terms of method and results. In this sense, and although it is unlikely that one model fits all country members, a first check is recommendable in relation to the model obtained previously. In case the bias achieved is too large, Steps 1-7 should be applied. Here, this process is carried out by employing the database from Honduras.

4. RESULTS AND DISCUSSION

In this Section, we present and discuss the results derived from the application of the methodology presented. Next, we analyse in detail the final network obtained for the cases of Nicaragua and Honduras. Finally, we conclude by presenting the limitations of the study.

4.1. Testing the proposed methodology

In this sub-section, main results are discussed as regards the application of the proposed methodology to two WaSH-related country databases (i.e. Nicaragua and Honduras). Initially, and according to the step-by-step proposal, different SLA and CIn tests are applied to the database of Nicaragua. A first scenario ("16 + 1") considers those sixteen components and the general index ("WSP") of the selected conceptual model. From the ten different primary networks obtained in first instance, six networks provided a structure containing a high amount of isolated nodes. Thus, only the remaining four networks are considered suitable for selection (see Appendix Figure A1). Considered the overall criteria proposed, the tandem "fast.iamb" SLA and "mi" CIn test conforms the selected primary network. Specifically, this tandem presents i) no isolated nodes, ii) 3 undirected links, and iii) 13 direct links pointing the objective node.

As explained in previous section, the first step involved in model calibration requires a manual refinement of the selected primary network. In doing so, links associated with the objective node are appropriately defined, undirected links are converted into direct ones and input nodes are established following the four proposed alternatives. As a result, four different networks are obtained (see Appendix Figure A2). Then, nodes CPTs are computed and evidences (IS distributions elicited from data) are assigned to input nodes. Finally, the inferred distributions associated with "WSP" node are compared to the ones provided by Nicaragua dataset. In this first scenario ("16 + 1"), acceptable biases were not achieved (see Appendix Table A5).

Thus, expert knowledge (in terms of structure) was provided to the candidate networks. In this case, expert knowledge is related to the provision of structure according to the selected conceptual model. A second scenario is tested by integrating "WSHL" and "WSSI" nodes, which represent the partial indices of the conceptual model. In this case, acceptable biases were obtained for those networks with "lowest IQR" and "smart" as inputs nodes. However, the test dataset was used to validate the networks and not acceptable biases were obtained (see Appendix Table A5).

In consequence, a third scenario was set out by including those nodes representing the four dimensions of the conceptual model ("WSL", "SHL", "WSI" and "SEP"). Again, four different networks are obtained (see Appendix Figure A6). Similarly, evidences are assigned to the input nodes and bias results are checked. In this case, the network where the input nodes are represented by those ones with lowest IQR provides slightly better results. Thus, this is the network selected. Then, the validation process is carried out by using the dataset kept for testing. Additionally, all dataset is used as well for the validation process (see Appendix Table A5). Additionally, further quality measures are checked.

Figure 3 summarises the step-by-step procedure described above, only focused on the selected network (fast.iamb + mi, lowest IQR as input nodes). First, it is seen how the biases are reduced due to structure provision of network structure and these fall within the threshold established (up to 5%). Specifically, when using the training dataset, this network infers a lower number of communities classified as "A" (3.7%) while overestimating the communities at state "B" (5.4%). On the other hand, those communities at state "C" are estimated with an insignificant bias. As a negative aspect, the model assigns a state "B" to the 1.5% of the communities (52 locations out of 3,495), when they have been classified as "D". Second, when applying the testing and the overall datasets, these biases are even lower.

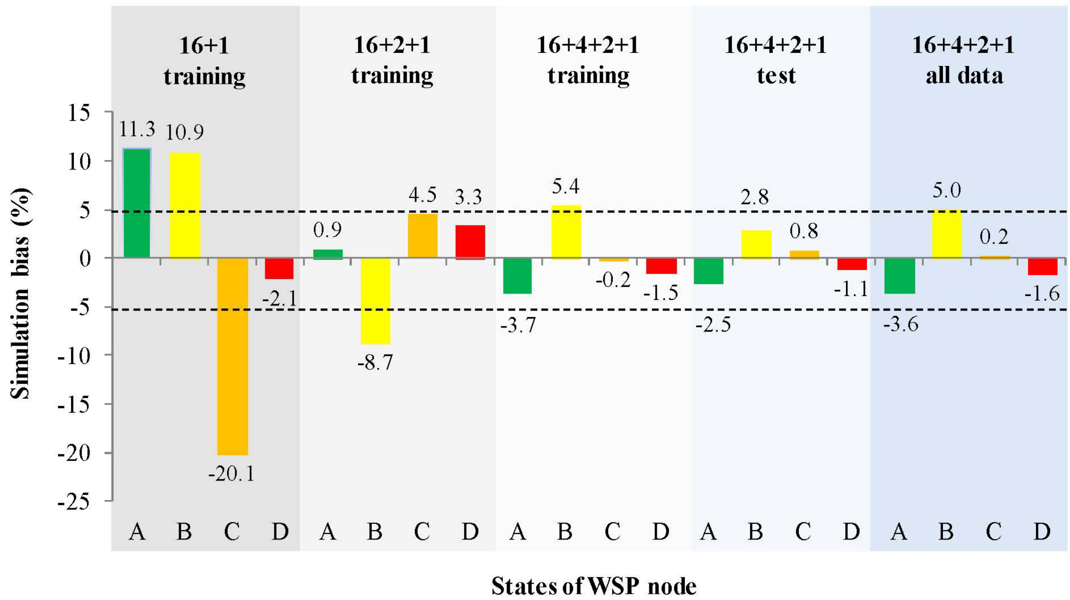


Figure 3. Resulting biases over WSP IS distribution associated to the stages of calibration and validation and for the case of Nicaragua. SLA + CIn test: fast.iamb (mi). Input nodes: lowest IQR. Results are obtained by subtracting the values provided by the database to those provided by the model (i.e. WSP_{model} - $WSP_{database}$).

On the other hand, Table 3 shows that, when focusing on the other variables of the network (i.e. dimensions and partial indices), the results provided by the model do not excess the threshold established. In consequence, the final network is positively validated.

Table 3. Further quality measures as to validate the network. Result biases are presented regarding the dimensions, partial indices and general index of the conceptual model. The full dataset of Nicaragua has been employed for this purpose.

Nodes	A	B	C	D
Wat. & San. Service Performance Index (WSP)	-3.6	5.0	0.2	-1.6
WaSH Service Level (WSHL)	-2.4	4.7	-1.6	-0.7
Water Services Sustainability Index (WSSI)	-4.4	4.2	1.6	-1.4
Water Service Level (WSL)	-0.8	2.3	-0.9	-0.6
Sanitation and Hygiene Service Level (SHL)	-3.2	3.3	4.0	-4.1
Water System Infrastructure (WSI)	-4.0	*5.5*	-0.6	-0.9
Service Provision (SEP)	-0.1	-1.6	2.9	-1.2

All values are express in percentage, taking as a reference the IS distributions of each node.
In Italics, those errors higher than 5%.

In summary, it has been seen how the biases are reduced due to structure provision. This is a major finding in terms of results. The introduction of intermediate variables makes conditional probability tables (CPTs) smaller and tractable (Marcot, 2017). This is coherent with the process carried out as the computational time is also reduced when providing structure to the network. However, we found out that introducing intermediate or summary variables (i.e. dimensions and partial indices) improves model inference capacity as well.

In order to validate the proposed methodology, in first instance, the final network obtained for Nicaragua was applied to the context of Honduras. In doing so, network input nodes were populated with the IS distributions associated with the context of the latter. Then, "WSP" inferred distributions were compared to the IS ones. In this case, results reached too large biases (e.g. up to 7% in WSP and up to 36% in WSI, see Appendix Table A6). However, the proposed step-by-step methodology was fully applied from the initial stage in order to check its validation. Similar to the previous BNs model, "fast.iamb (mi)" was the final selection of SLA and CIn test. In addition to this, lower biases were obtained as well when providing full structure (expert knowledge), but in this case under a "smart variables" approach (see Appendix Table A7).

Table 4. Result biases regarding the dimensions, partial indices and general index of the conceptual model. The full dataset of Honduras has been used in this case.

Nodes	A	B	C	D
Wat. & San. Service Performance Index (WSP)	-3.6	*6.3*	-2.3	-0.4
WaSH Service Level (WSHL)	-1.8	3.9	-1.8	-0.3
Water Services Sustainability Index (WSSI)	*-5.1*	*7.9*	-2.0	-0.8
Water Service Level (WSL)	-0.7	1.6	-0.8	-0.1
Sanitation and Hygiene Service Level (SHL)	-2.6	*5.1*	-0.8	-1.7
Water System Infrastructure (WSI)	-1.1	1.5	0.5	-0.9
Service Provision (SEP)	*-5.9*	*9.1*	-1.9	-1.3

All values are express in percentage, taking as a reference the IS distributions of each node.
In Italics, those errors higher than 5%.

In Table 4, final results (in terms of errors) are provided. It can be seen that, in six cases, the biases reach a value up to 9.1%. Even if the desired goodness of the model was not fully achieved, it is considered that the overall results are positive enough to not discard it (79% of the values inferred are below the threshold established). However, two measures are proposed to tackle this situation. First, the initial threshold of 5% might be redefined according to the problem at hand and assessed whether or not it is too strict. Second, and in order to counteract this systematic error, it is proposed a simple action when new evidences (new "A-B-C-D") distributions are provided to the network input nodes. Considering the IS distributions as "x" and the inferred values as "y", the difference "x - y" (bias) is obtained (as shown in Table 4). When new evidences are provided, new inferred values "z" are obtained. In this way, a correction of the results should follow the operation "z - (x - y)". Having said this, there is insight to confirm the validity of the proposed methodology for network construction regardless the context at hand.

4.1. The network of Nicaragua

The graphical result obtained through the application of the proposed methodology is presented in Figure 4. First, we highlight the reduction in the number of input variables which explain the general index “WSP”. Thus, the BNs model identifies, as key components, those ones related to “water accessibility” (ACC), “seasonality” (SEA), “sanitation service level” (SSL), “community hygiene” (COM), “water system infrastructure” (INF), “water catchment protection” (PRO) and service provider’s “operation and maintenance (OPM) and environmental (ENV) management”. This aspect is especially relevant for data updating purposes, as the questionnaires which feed the conceptual framework collect an important amount of information. In this sense, it is possible to differentiate two data sets. First, that one collected only once and which are not considered in the conceptual model (e.g. name of the community, administrative scales, geospatial coordinates of water system elements, etc.). Second, those data integrated in the conceptual model. Specifically, this model is fed with a set of 43 questions. If the case of Nicaragua is considered, the 8 input nodes obtained would just require 22 questions (51%) to infer the values of the composite indicators and aggregated indices. In consequence, important savings in terms of time (and so economic ones) are achieved.

Second, we also remark the identification of links among the different components at hand, which appear as an advantage of the BNs approach in relation to CI-based one. In this sense, it must be reminded that these links represent dependencies rather than cause-effect linkages. On the one hand, it is possible to identify intra-components relationships. Here, the model identifies coherent dependencies among hygiene-related components (i.e. community (COM), household (WAT) and personal (PER)). Equally coherent are those links between water “seasonality” (SEA) and “water quality” (QUA), and “water quality” and service “continuity” (CON), which have been already documented. The former argues the possibility of accessing higher risk water sources via changes in the type of primary water source used by households during the dry season (Pearson et al., 2016). The latter points out that providing water intermittently can compromise water quality in the distribution system (Kumpel and Nelson, 2016). Deepening into Nicaragua database (i.e. checking specific contingency tables), it is observed that those communities lacking of seasonality problems (i.e. qualified as “A”), lack as well of water quality issues (43% of the rural communities). The same statement can be done when observing those communities evaluated as “A” as regards QUA and CON (38% of the rural communities). In addition to this, the model identifies a dependency between the status of the “production infrastructure” (INF) and the “treatment system” (TRE). As far as the service provider performance, it is seen that the four components are connected among them. On the other hand, the model only identifies in this case one inter-component relationship (i.e. the link between service provider "economic management" (ECO) and "system autonomy" (AUT)). From Nicaragua database, it is observed a higher correlation when ECO and AUT are evaluated as "D" (20% of the rural communities). AUT is evaluated as the capacity of the storage system (Pérez-Foguet and Flores-Baquero, 2015) and it seems coherent that the lack of economic resources is related to the construction of appropriate storage infrastructures.

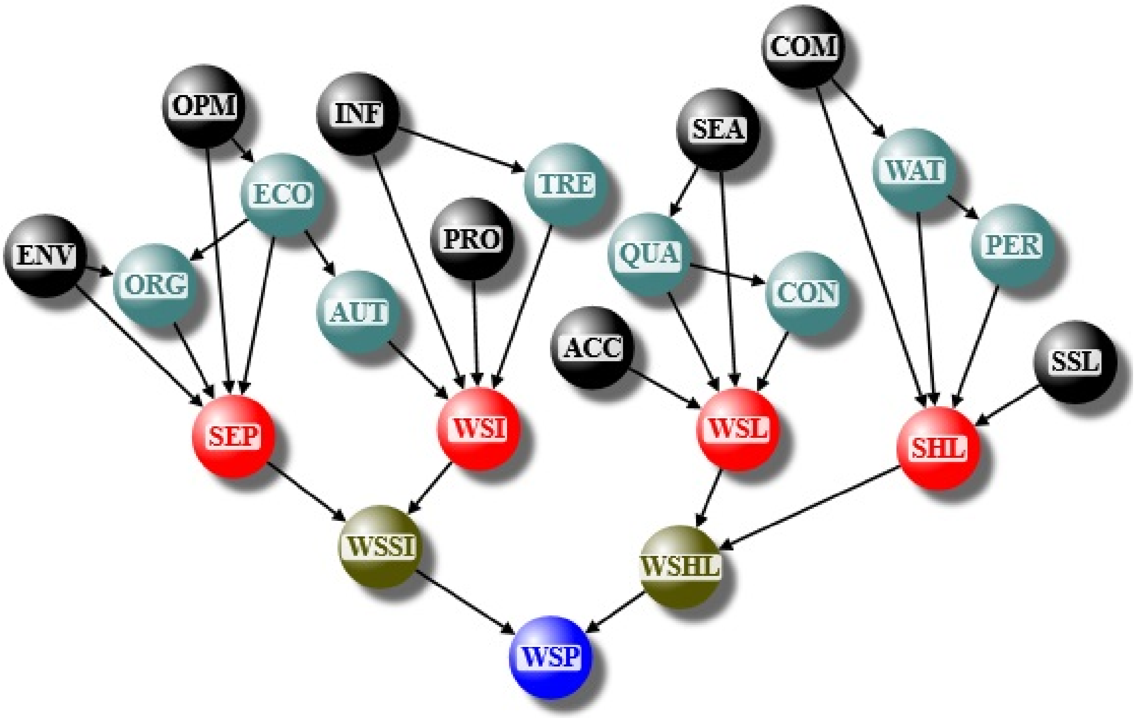


Figure 4. Final network for the case of Nicaragua (software NodeXL has been used to depict the results). In black, input nodes. In grey, intermediate nodes in terms of components. In red, nodes representing conceptual model dimensions. In dark green, nodes representing partial indices. In blue, WSP objective node.

A sensitivity analysis was carried out to identify, among the input nodes, the key contributors on water and sanitation performance index for rural communities (WSP). In doing so, we applied an inverse use of the network, as recently tested by Li et al. (2019). Also known as diagnostic inference (Carriger et al., 2016), the model is run in a backward direction (from objective to input nodes). Specifically, Figure 5 shows the results of predicted changes of all input variables when assuming a hard evidence as far as a sustainable WaSH service provision (i.e. WSP = "A"). The sensitivity analysis suggests that "operation and maintenance management" (OPM) and "community hygiene" (COM) are the two most important components affecting WSP index. These are followed by "sanitation service level" (SSL), "water system infrastructure" (INF), and "seasonality" (SEA). Those components related to "environmental management" (ENV), "water catchment protection" (PRO) and "water accessibility" (ACC) appeared to have the least impact. If the aim is to achieve sustainable services and to support those communities more in need, great efforts might be oriented to those communities with a grade “D” in relation to OPM (16%), SSL (12%), COM (11%) and SEA (7%).

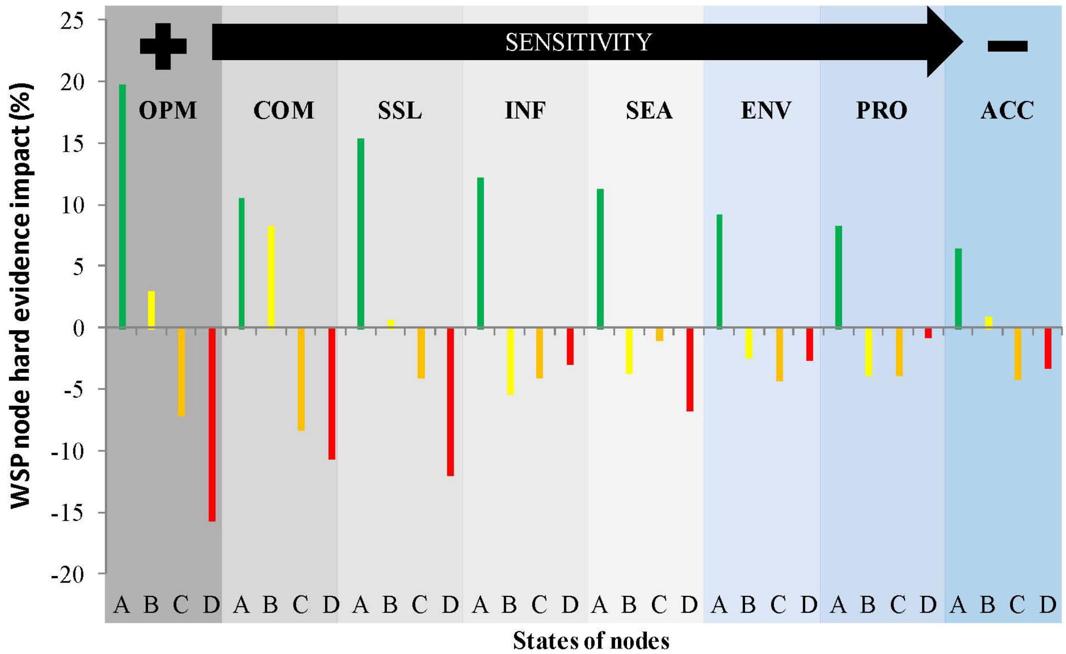


Figure 5. Nicaragua sensitivity analysis results and predicated impacts on inputs states with the objective node of WSP set at A=100%.

4.3. The network of Honduras

The network obtained for the case of Honduras is depicted in Figure 6. Similar to the previous result, the number of components explaining “WSP” node are reduced to the half. Although similar key variables are identified by the BNs model (i.e. ACC, SSL and PRO), the results present much more differences. On the one hand, the key explanatory components rely on “continuity” (CON), “personal hygiene” (PER), “water system treatment” (TRE) and service provider’s “organizational” (ORG) and “economical” (ECO) management. On the other hand, a higher number of inter-component relationships are identified.

First, we highlight the fact that the input nodes of this network are so-called “smart”. In this case, the time saving in data updated would be even greater comparing to Nicaragua. Remarkably, it wouldn’t be necessary to visit the different elements of the water system, which normally is a time consuming task, mostly in disperse rural areas.

Second, it is important to highlight as well the existing differences when the model identifies links among components. However, we recall to the data-driven nature of this research. In this case, we find coherent as well the results obtained. For example, and focusing on inter-component relationships, “system treatment” (TRE) and “water quality” (QUA) are logical connections. Similarly, “water catchment protection” (PRO) is associated with the "environmental management" (ENV) of the service provider. The model also identifies a dependency between "seasonality" (SEA) and "system autonomy" (AUT). Contingency tables from Honduras database shows that in 37% of the rural communities a grade of "A" for these two components is achieved. Similarly, in 43% of the communities the best qualification is achieved simultaneously for service provider's ENV and the status of the "production infrastructure" (INF). Surprisingly, the model identifies the link between "household hygiene" (WAT) and "operation and maintenance management" (OPM). As far as intra-component linkages, coherent results are found in relation to those dependencies between "sanitation service level" (SSL) and "community hygiene" (COM), and "seasonality" (SEA) with "accessibility" (ACC) and "continuity" (CON). In contrast, the model does not identify any dependency among the components associated to the service provider performance.

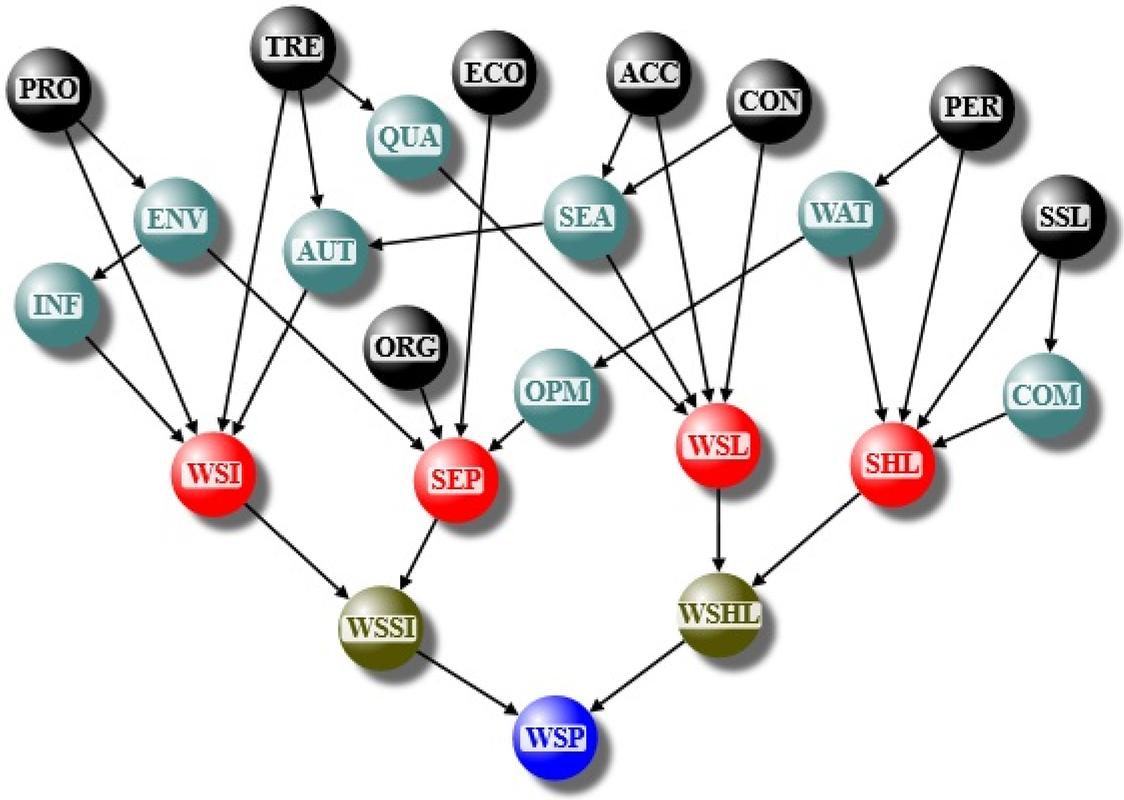


Figure 6. Honduras final network. In black, input nodes. In grey, intermediate nodes in terms of components. In red, conceptual model dimensions. In dark green, partial indices. In blue, WSP general index.

Similarly to Nicaragua, we performed a sensitivity analysis following the same procedure described previously (see Figure 7). In this case, the analysis suggests that "treatment system" (TRE) and "personal hygiene" (PER) are the two most influencing components on the WSP index. These are followed by "water catchment protection" (PRO), "sanitation service level"

(SSL) and "continuity" (CON) and, ultimately, "economic" (ECO) and "organizational" (ORG) management and "accessibility" (ACC).

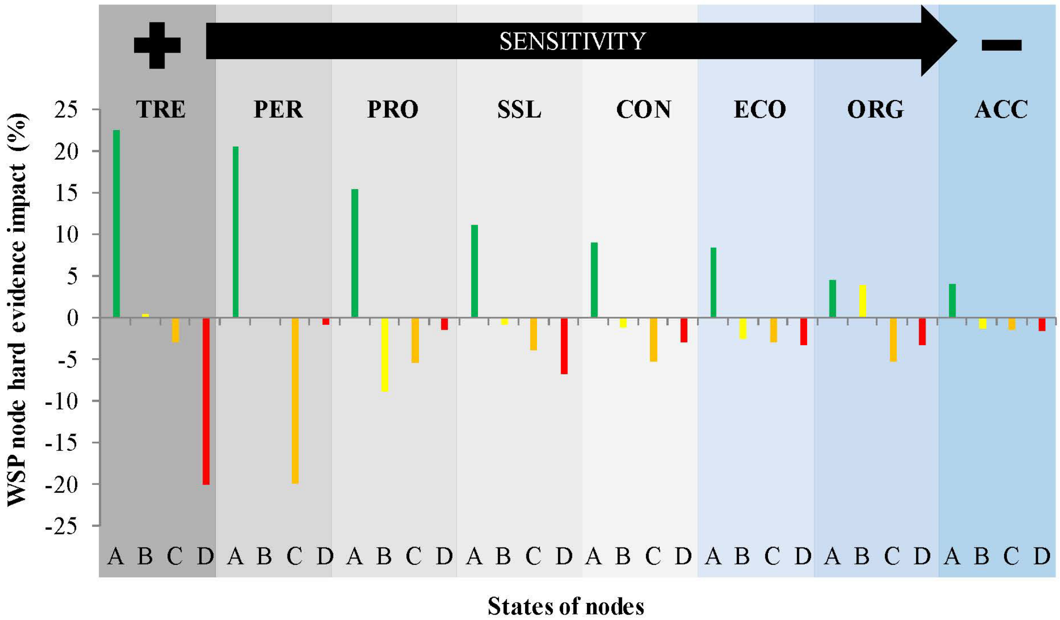


Figure 7. Honduras sensitivity analysis results and predicated impacts on inputs states with the objective node of WSP set at A=100%.

4.4. Limitations of the study

Due to data availability, we have not reproduced the analysis in other contexts, which it is of interest to provide a more general methodology. Similarly, it was not applied where there is a much lower amount of data, as this fact hinders the definition of an initial DAG (Step A.1) and, thus, the application of the proposed methodology. However, score-based algorithms are presented as an alternative to test for network generation.

We have not extended this research to compare the results of using, for example, a higher number of variable states or continuous values. We are concerned that this aspect might be sensitive to the structure learning process (Alameddine et al., 2011). In addition, this fact, while adequately facilitating node CPTs management, represents a potential loss of statistical accuracy (Chen and Pollino, 2012).

4. CONCLUSIONS

In this study, we propose a step-by-step methodology to reproduce composite indicators-based conceptual frameworks which integrate a hierarchical structure. In doing so, the flexibility of Bayesian Networks has been exploited. The case of a regional information system has been selected as a pilot study. The methodology has been successfully calibrated and validated. As a result, we have developed a semi-automatic procedure (manual decisions in one step of the method), which relies basically on a data-driven approach.

Obtained results have shown how expert knowledge (in terms of network structure) improves model inference capacity. The biases achieved for the case of Nicaragua do not reach the established threshold of 5%. Similarly, positive results were obtained for the case of Honduras. However, in 6 cases out of 28 (12%), the biases reach a value up to 9.1%. Even if the desired goodness of the model was not fully achieved, it is considered that the overall results are positive enough to not discard it.

Additionally, the two models obtained allowed us to identify the key variables that explain the objective one (in this case, the general index of the selected conceptual framework). This fact is especially important when considering the existing difficulties of the sector in updating the required data. In consequence, important savings in terms of time (and so economic ones) can be achieved. In addition to this, the flexibility of BNs permitted us to quantify, for each context, the key contributors on the water and sanitation performance index for rural communities.

BNs are able to identify as well the interdependencies of variables at hand. This fact makes this approach more suitable than the use of composite indicators. In this sense, it might enhance multi- and trans-disciplinary actions.

Extending the proposal to address the limitations presented in this study, as well as working directly with the raw variables integrated in the selected conceptual framework, suggests the way forward.

REFERENCES

Aguilera, P.A., Fernández, A., Fernández, R., Rumí, R., Salmerón, A., 2011. Bayesian Networks in Environmental Modelling. Environ. Modell. Softw. 26 (12), 1376–1388. https://doi.org/10.1016/j.envsoft.2011.06.004.

Aguilera, P.A., Fernández, A., Ropero, R.F., Molina, L., 2013. Groundwater Quality Assessment Using Data Clustering Based on Hybrid Bayesian Networks. Stoch. Environ. Res. Risk Assess. 27 (2), 435–447. https://doi.org/10.1007/s00477-012-0676-8.

Alameddine, I., Cha, Y.K., Reckhow, K.H., 2011. An Evaluation of Automated Structure Learning with Bayesian Networks: An Application to Estuarine Chlorophyll Dynamics. Environ. Modell. Softw. 26 (2), 163–172. https://doi.org/10.1016/j.envsoft.2010.08.007.

Bandura, R., 2006. A Survey of Composite Indices Measuring Country Performance: 2006 Update. United Nations Development Programme – Office of Development Studies.

Bordalo, A.A., Teixeira, R., Wiebe, W. J., 2006. A Water Quality Index Applied to an International Shared River Basin: The Case of the Douro River. Environ. Manage. 38 (6), 910–920. https://doi.org/10.1007/s00267-004-0037-6.

Bromley, J., 2005. Guidelines for the Use of Bayesian Networks as a Participatory Tool for Water Resource. Walliford, United Kingdom.

Carriger, J.F., Barron, M.G., Newman, M.C., 2016. Bayesian Networks Improve Causal Environmental Assessments for Evidence-Based Policy. Environ. Sci. Technol. 50 (24), 13195–13205. https://doi.org/10.1021/acs.est.6b03220.

Chandrakumar, C., McLaren, S.J., 2018. Towards a Comprehensive Absolute Sustainability Assessment Method for Effective Earth System Governance: Defining Key Environmental Indicators Using an Enhanced-DPSIR Framework. Ecol. Indic. 90, 577–583. https://doi.org/10.1016/j.ecolind.2018.03.063.

Chen, S.H., Pollino, C.A., 2012., Good Practice in Bayesian Network Modelling. Environ. Modell. Softw., 37, 134–145. https://doi.org/10.1016/j.envsoft.2012.03.012.

Cho, D.I., Ogwang, T., Opio, C., 2010. Simplifying the Water Poverty Index. Soc. Indic. Res. 97 (2), 257–267. https://doi.org/10.1007/s11205-009-9501-2.

Chung, E.S., Lee, K.S., 2009. Prioritization of Water Management for Sustainability Using Hydrologic Simulation Model and Multicriteria Decision Making Techniques. Environ. Manage. 90 (3), 1502–1511. https://doi.org/10.1016/j.jenvman.2008.10.008.

Cohen, A., Sullivan, C.A., 2010. Water and Poverty in Rural China: Developing an Instrument to Assess the Multiple Dimensions of Water and Poverty. Ecol. Econ. 69 (5), 999–1009. https://doi.org/10.1016/j.ecolecon.2010.01.004.

Cronk, R., Bartram, J., 2017. Factors Influencing Water System Functionality in Nigeria and Tanzania: A Regression and Bayesian Network Analysis. Environ. Sci. Technol. 51 (19), 11336–11345. https://doi.org/10.1021/acs.est.7b03287.

Cronk, R., Bartram, J., 2018. Identifying Opportunities to Improve Piped Water Continuity and Water System Monitoring in Honduras, Nicaragua, and Panama: Evidence from Bayesian Networks and Regression Analysis. J. Clean. Prod. 196, 1–10. https://doi.org/10.1016/j.jclepro.2018.06.017.

Dondeynaz, C., López Puga, J., Carmona Moreno, C., 2013. Bayesian Networks Modelling in Support to Cross-cutting Analysis of Water Supply and Sanitation in Developing Countries. Hydrol. Earth Syst. Sci. 17 (9), 3397–3419. https://doi.org/10.5194/hess-17-3397-2013.

EEA (European Environment Agency), 1999. Environmental Indicators : Typology and Overview. Technical report nº 25. Copenhagen, Denmark.

Fisher, M.B., Shields, K.F., Chan, T.U., Christenson, E., Cronk, R.D., Leker, H., Samani, D., Apoya, P., Lutz, A., Bartram, J., 2015. Understanding Handpump Sustainability: Determinants of Rural Water Source Functionality in the Greater Afram Plains Region of Ghana. Water Resour. Res. 51 (10), 8431–8449. https://doi.org/10.1002/2014WR016770.

Flores-Baquero, O., Gallego Ayala, J., Giné-Garriga, R., Jiménez, A., Pérez-Foguet, A., 2016. The influence of the Human Rights to Water and Sanitation normative content in measuring the level of service. Soc. Indic. Res. 133 (2), 763–786. https://doi.org/10.1007/s11205-016-1374-6.

Garcia-Prats, A., González-Sanchis, M., Del Campo, A. D., Lull, C., 2018. Hydrology-oriented Forest Management Trade-offs. A Modeling Framework Coupling Field Data, Simulation Results and Bayesian Networks. Sci. Total Environ. 639, 725–741. https://doi.org/10.1016/j.scitotenv.2018.05.134.

García-Santos, G., Madruga De Brito, M., Höllermann, B., Taft, L., Almoradie, A., Evers, M., 2018. Methodology to Explore Emergent Behaviours of the Interactions between Water Resources and Ecosystem under a Pluralistic Approach. P. Int. Ass. Hydrol. Sci. 379, 83–87. https://doi.org/10.5194/piahs-379-83-2018.

Giné-Garriga, R., Flores-Baquero, O., Jiménez-Fdez. de Palencia, A., Pérez-Foguet, A. 2017. Monitoring sanitation and hygiene in the 2030 Agenda for Sustainable Development: A review through the lens of human rights. Sci. Total Environ. 580, 1108–1119. https://doi.org/10.1016/j.scitotenv.2016.12.066.

Giné-Garriga, R., Pérez-Foguet, A., 2010. Improved Method to Calculate a Water Poverty Index at Local Scale. J. Environ. Eng. 136, 1287–1298. https://doi.org/10.1061/(ASCE)EE.1943-7870.0000255.

Giné-Garriga, R., Pérez-Foguet, A., 2011. Application of a Revised Water Poverty Index to Target the Water Poor. Water Sci. Technol. 63 (6), 1099–1110. https://doi.org/10.2166/wst.2011.347.

Giné-Garriga, R., Pérez-Foguet, A., 2013. Unravelling the Linkages between Water, Sanitation, Hygiene and Rural Poverty: The WASH Poverty Index. Water Resour. Manage. 27 (5), 1501–1515. https://doi.org/10.1007/s11269-012-0251-6.

Giné-Garriga, R. Pérez-Foguet, A., 2018. 'Measuring sanitation poverty: a multidimensional measure to assess delivery of sanitation and hygiene services at the household level'. OPHI Working Paper 116, University of Oxford. Available at https://ophi.org.uk/measuring-sanitation-poverty-a-multidimensional-measure-to-assess-delivery-of-sanitation-and-hygiene-services-at-the-household-level/)

Giné-Garriga, R., Requejo-Castro, D., Molina, J.L., Pérez-Foguet, A., 2018. A Novel Planning Approach for the Water, Sanitation and Hygiene (WaSH) sector: The Use of Object-oriented Bayesian Networks. Environ. Modell. Softw. 103, 1–15. https://doi.org/10.1016/j.envsoft.2018.01.021.

Godfrey, L., Tood, C., 2001. Defining Thresholds for Freshwater Sustainability Indicators within the Context of South African Water Resource Management. In 2nd WARFSA / WaterNet Symposium : Integrated Water Resource Management : Theory , Practice , Cases. Cape Town, South Africa.

Godfrey, S., van der Velden, M., Muianga, A., Xavier, A., Downs, K., Morgan, C., Bartram, J., 2014. Sustainability Check: Five-year Annual Sustainability Audits of the Water Supply and Open Defecation Free Status in the 'One Million Initiative', Mozambique. J. Water Sanit. Hyg. Dev. 4 (3), 471–483. https://doi.org/10.2166/washdev.2014.118.

Hou, Y., Zhou, S., Burkhard, B., Müller, F., 2014. Socioeconomic Influences on Biodiversity, Ecosystem Services and Human Well-being: A Quantitative Application of the DPSIR Model in Jiangsu, China. Sci. Total Environ. 490, 1012–1028. https://doi.org/10.1016/j.scitotenv.2014.05.071.

Joint Monitoring Programme, 2000. Global Water Supply and Sanitation Assessment 2000 Report. Water Supply. WHO/UNICEF. Geneva, Switzerland / New York, United States.

Joint Monitoring Programme, 2008. Progress on Drinking Water and Sanitation: Special Focus on Sanitation. Geneve, Switzerland.

Joint Monitoring Programme, 2017. Progress on Drinking Water , Sanitation and Hygiene: 2017 Update and SDG Baselines. Geneve, Switzerland.

Kumar, A., Mazumdar, K., 2002. Use of Bayesian Networks for Monitoring of Total Sanitation Campaign Projects-Indian. In 28th WEDC Conference on Sustainable Environmental Sanitation and Water Services. Kolkata, India.

Kumpel, E., Nelson, K.L., 2016. Intermittent Water Supply: Prevalence, Practice, and Microbial Water Quality. Environ. Sci. Technol. 50 (2), 542–553. https://doi.org/10.1021/acs.est.5b03973.

Li, J., Song, G., Semakula, H.M., Zhang, Sh., 2019. Climatic Burden of Eating at Home against Away-from-home: A novel Bayesian Belief Network Model for the Mechanism of Eating-out in Urban China. Sci. Total Environ. 650, 224–232. https://doi.org/10.1016/j.scitotenv.2018.09.015.

Lin, T., Lin, J., Cui, S., Cameron, S. 2009. Using a Network Framework to Quantitatively Select Ecological Indicators. Ecol. Indic. 9 (6), 1114–1120. https://doi.org/10.1016/j.ecolind.2008.12.009.

Liu, H., Zhou, S., Lam, W., & Guan, J. 2017. A New Hybrid Method for Learning Bayesian Networks: Separation and Reunion. Knowl-Based Syst. 121, 185–197. https://doi.org/10.1016/j.knosys.2017.01.029.

Luh, J., Baum, R., Bartram, J., 2013. Equity in Water and Sanitation: Developing an Index to Measure Progressive Realization of the Human Right. Int. J. Hyg. Environ. Heal. 216 (6), 662–671. https://doi.org/10.1016/j.ijheh.2012.12.007.

Madsen, A.L., Jensen, F., Salmerón, A., Langseth, H., Nielsen, T.D., 2016. A Parallel Algorithm for Bayesian Network Structure Learning from Large Data Sets. Knowl.-Based Syst. 117, 46–55. https://doi.org/10.1016/j.knosys.2016.07.031.

Marcot, B.G. 2017. Common Quandaries and their Practical Solutions in Bayesian Network Modeling. Ecol. Model. 358, 1–9. https://doi.org/10.1016/j.ecolmodel.2017.05.011.

Nardo, M., Saisana, M., Saltelli, A., Tarantola, S., Hoffman, A., Giovannini, E., 2005. Handbook on Constructing Composite Indicators: Methodology and User Guide. OECD Statistics Working Paper. OECD, Paris.

Niemeijer, D., De Groot, R.S., 2008. Framing Environmental Indicators: Moving from Causal Chains to Causal Networks. Environ. Dev. Sustain. 10 (1), 89–106. https://doi.org/10.1007/s10668-006-9040-9.

OCDE., 1993. OECD Core Set of Indicators for Environmental Performance Reviews: A Synthesis Report by the Group on the State of the Enviroment. OECD Environmental Directorate Monographs nº 83. Paris, France.

Ortiz-Lozano, L, 2012. Identification of Priority Conservation Actions in Marine Protected Areas: Using a Causal Networks Approach. Ocean Coast. Manage. 55, 74–83. https://doi.org/10.1016/j.ocecoaman.2011.10.013.

Pearson, A.L., Zwickle, A., Namanya, J., Rzotkiewicz, A., Mwita, E., 2016. Seasonal Shifts in Primary Water Source Type: A Comparison of Largely Pastoral Communities in Uganda and Tanzania. Int. J. Env. Res. Pub. He. 13 (2), 169. https://doi.org/10.3390/ijerph13020169.

Pérez-Foguet, A., Flores-Baquero, O., 2015. Decision Support Model for SIASAR. Technical Proposal (v. 2015.07). Universitat Politècnica de Catalunya, Barcelona. Available at: http://hdl.handle.net/2117/77587.

Pérez-Foguet, A., Giné-Garriga, R., 2011. Analyzing Water Poverty in Basins. Water Resour. Manage. 25 (14), 3595–3612. https://doi.org/10.1007/s11269-011-9872-4.

Phan, Th.D., Smart, C.R.J., Capon, S.J., Hadwen, W.L., Sahin, O., 2016. Applications of Bayesian belief networks in water resource management: A systematic review. Environ. Model. Softw. 85, 98–111. http://dx.doi.org/10.1016/j.envsoft.2016.08.006.

Pirrone, N., Trombino, G., Cinnirella, S., Algieri, A., Bendoricchio, G., Palmeri, L., 2005. The Driver-Pressure-State-Impact-Response (DPSIR) Approach for Integrated Catchment-coastal Zone Management: Preliminary Application to the Po Catchment-Adriatic Sea Coastal Zone System. Reg. Environ. Change 5 (2–3), 111–137. https://doi.org/10.1007/s10113-004-0092-9.

Ramos-Quintana, F., Ortíz-Hernández, M.L., Sánchez-Salinas, E., Úrsula-Vázquez, E., Guerrero, J.A., Zamorano, M., 2018. Quantitative-qualitative Assessments of Environmental Causal Networks to Support the DPSIR Framework in the Decision-making Process. Environ. Impact Asses. 69, 42–60. https://doi.org/10.1016/j.eiar.2017.11.004.

Requejo-Castro, D., Giné-Garriga, R., Flores-Baquero, O., Martínez, G., Rodríguez, A., Jimenez-Fdez. de Palencia, A., Pérez-Foguet, A., 2017. SIASAR: A Country-led Indicator Framework for Monitoring the Rural Water and Sanitation Sector in Latin America and the Caribbean. Water Pract. Technol. 12 (2), 372–385. https://doi.org/10.2166/wpt.2017.041.

Saaty, T.L., 2001. Decision making with dependence and feedback: the Analytic Network Process. RWS Publishing. University of Pittsburgh, Pennsylvania, United States.

Scharin, H., Ericsdotter, S., Elliott, M., Turner, R.K., Niiranen, S., Blenckner, T., Hyytiäinen, K., Ahlvik, L., Ahtiainen, H., Artell, J., Hasselström, L., Söderqvist, T., Rockström, J., 2016. Processes for the Sustainable Stewardship of Marine Environments. Ecol. Econ. 128, 55–67. https://doi.org/10.1016/j.ecolecon.2016.04.010.

Scutari, M., 2010. Learning Bayesian Networks with the bnlearn R Package. J. Stat. Softw. 35 (3), 1–22. https://doi.org/10.18637/jss.v035.i03.

Scutari, M., Denis, J.B., 2015. Bayesian Networks with Examples in R. Taylor & Francis Group. Boca Raton, Florida. ISBN: 978-1-4822-2559-4.

SIASAR, 2017. Field Data Collection Protocol v10 (Protocolo de campo v10): SIASAR 2.0. Available at: http://www.siasar.org/es/content/documentacion-tecnica.

SIASAR, 2018. Information Validation Protocol v2 (Protocolo de Validación de Información v2). Available at: http://www.siasar.org/es/content/colombia.

Singh, R.K., Murty, H.R., Gupta, S.K., Dikshit, A.K., 2009. An Overview of Sustainability Assessment Methodologies. Ecol. Indic. 9 (2), 189–212. https://doi.org/10.1016/j.ecolind.2008.05.011.

Song, G., Semakula, H.M., Fullana-i-Palmer, P., 2018. Chinese Household Food Waste and its' Climatic Burden Driven by Urbanization: A Bayesian Belief Network Modelling for Reduction Possibilities in the Context of Global Efforts. J. Clean. Prod. 202, 916–924. https://doi.org/10.1016/j.jclepro.2018.08.233.

Sullivan, C.A., Meigh, J.R., Giacomello, A.M., Fediw, T., Lawrence, P., Samad, M., Mlote, S., Hutton, C., Allan, J.A., Schulze, R.E., Dlamini, D.J., Cosgrove, W., Delli Priscoli, J., Gleick, P., Smout, I., Cobbing, J., Calow, R., Hunt, C., Hussain, A., Acreman, M.C., King, J., Malomo, S., Tate, E.L., O'Regan, D., Milner, S. Steyl, I., 2003. The Water Poverty Index: Development and Application at the Community Scale. Nat. Resour. Forum 27 (3), 189–199. https://doi.org/10.1111/1477-8947.00054.

Taft, L., Evers, M., 2016. A Review of Current and Possible Future Human-water Dynamics in Myanmar's River Basins. Hydrol. Earth Syst. Sci., 20 (12), 4913–4928. https://doi.org/10.5194/hess-20-4913-2016.

UN (United Nations), 1996. Indicators of sustainable development framework and methodologies. United Nations Sales Publication No E 96 II A 16. New York, United States.

Vacik, H., Wolfslehner, B., Seidl, R., Lexer, M.J., 2007. Integrating the DPSIR-approach and the Analytic Network Process for the Assessment of Forest Management Strategies. In: Reynolds, K., Rennolls, K., Kohl, M., Thomson, A., Shannon, M., Ray, D. (Eds.), Sustainable Forestry: From Monitoring and Modelling to Knowledge Management and Policy Science. CABI Publishing. Oxfordshire, United Kingdom. https://doi.org/10.1079/9781845931742.0393.

Webb, A.L., Stein, A. D., Ramakrishnan, U., Hertzberg, V.S., Urizar, M., Martorell, R.. 2006. A Simple Index to Measure Hygiene Behaviours. Int. J. Epidemiol. 35 (6), 1469–1477. https://doi.org/10.1093/ije/dyl165.

Wolfslehner, B., Vacik, H., 2011. Mapping Indicator Models: From Intuitive Problem Structuring to Quantified Decision-making in Sustainable Forest Management. Ecol. Indic. 11 (2), 274–283. https://doi.org/10.1016/j.ecolind.2010.05.004.

WSP (World Bank Water and Sanitation Programme). 2015. Developing Business Models for Fecal Sludge Management in Maputo. Maputo, Mozambique.

Table A1. Dimensions, components and indicators (including corresponding survey questionnaire) within SIASAR conceptual framework. Source: Pérez-Foguet and Flores-Baquero (2015).

Dimensions	Components	Indicators
Water service level (WSL)	Accessibility (ACC)	Improved water supply coverage [a] Access time [b]
	Continuity (CON)	Service hours per day [b]
	Seasonality (SEA)	Water system flow [b] Minimum water system flow [b] Sufficient water during summer [b] Number of households [a]
	Quality (QUA)	Physico-chemical quality [b] Bacteriological quality [b]
Sanitation and hygiene service level (SHL)	Sanitation service level (SSL)	Improved sanitation coverage [a]
	Personal hygiene (PER)	Hand-washing practice [a]
	Household hygiene (WAT)	Household safety water management [a]
	Community hygiene (COM)	Garbage and puddles presence [a] Open defecation [a]
Water system infrastructure (WSI)	System autonomy (AUT)	Service days without production [b] Number of households [a]
	Production infrastructure (INF)	Catchment area status [b] Conduction status [b] Storage status [b] Distribution status [b]
	Water catchment area protection (PRO)	Catchment protection area status [b]
	Treatment system (TRE)	Treatment system typology [b] Treatment system functionality [b]
Service provision (SEP)	Organization management (ORG)	Legalization and directive structure [c] Ordinary operation [c] Equity within the organization [c] Economic management and accountability [c]
	Operation & maintenance management (OPM)	O&M general assessment [c] Basic operation with chlorine [b]
	Economic management (ECO)	Collection efficiency rate [c] Cost coverage rate [c] Liquid assets rate [c]
	Environmental management (ENV)	Catchment attention measurements [c] Environmental sanitation promotion [c]

(a) Community questionnaire, (b) System questionnaire, and (c) Service provision questionnaire.

Table A2. Aggregation to construct the different dimensions, partial indices and general index.

Water and sanitation performance index for rural communities (WSP)

$$WSP = \prod_{i,j=0}^{i,j=1} x_j^{\,p_j} = (WSHL * WSSI)^{1/2}$$

WaSH service level index (WSHL)

$$WSHL = \prod_{i,j=0}^{i,j=1} x_j^{\,p_j} = (WSL * SHL)^{1/2}$$

Water service level (WSL)

$$WSL = \sum_{i,j=0}^{i,j=1} x_i \cdot p_j = \frac{(ACC + CON + SEA + QUA)}{4}$$

ACC - Accessibility

CON - Continuity

SEA - Seasonality

QUA – Quality

Sanitation and hygiene service level (SHL)

$$SHL = \sum_{i,j=0}^{i,j=1} x_i \cdot p_j = \frac{(SSL + PER + WAT + COM)}{4}$$

SSL - Sanitation service level

PER - Personal hygiene

WAT - Household hygiene

COM - Community hygiene

Water services sustainability index (WSSI)

$$WSSI = \prod_{i,j=0}^{i,j=1} x_j^{\,p_j} = (WSI * SEP)^{1/2}$$

Water system infrastructure (WSI)

$$WSI = \sum_{i,j=0}^{i,j=1} x_i \cdot p_j = \frac{(AUT + INF + PRO + TRE)}{4}$$

AUT - System autonomy

INF - Production infrastructure

PRO - Water catchment area protection

TRE - Treatment system

Service provision - SEP

$$SEP = \sum_{i,j=0}^{i,j=1} x_i \cdot p_j = \frac{(ORG + OPM + ECO + ENV)}{4}$$

ORG - Organization management

OPM - Operation & maintenance management

ECO - Economic management

ENV - Environmental management

Table A3. Variables distributions detailed numerically.

	NICARAGUA				HONDURAS			
	States (%)				States (%)			
Variables	**A**	**B**	**C**	**D**	**A**	**B**	**C**	**D**
Accessibility	23.1	41.5	21.4	14.0	42.1	44.2	10.2	3.5
Continuity	56.9	14.5	11.7	16.9	79.5	4.4	10.8	5.3
Seasonality	64.7	19.1	4.7	11.5	62.4	25.8	4.5	7.3
Quality	63.3	6.9	1.2	28.6	12.8	6.8	0.2	80.2
Sanitation service level	36.7	23.1	14.9	25.3	42.8	29.0	13.8	14.4
Personal hygiene	30.4	---	66.9	2.7	55.3	---	43.2	1.5
Household hygiene	39.2	---	47.1	13.7	28.2	---	46.3	25.5
Community hygiene	10.1	30.0	36.7	23.2	20.8	30.1	37.2	11.9
System autonomy	25.4	15.6	20.2	38.8	54.6	18.0	11.5	15.9
Production infrastructure	54.5	32.6	7.8	5.1	61.9	29.2	4.2	4.7
Water catchment protection	45.1	36.9	15.9	2.1	48.7	39.2	9.8	2.3
Treatment system	53.7	6.5	9.9	29.9	23.7	1.0	30.5	44.8
Organizational management	12.0	24.3	26.4	37.3	15.9	42.0	31.1	11.0
O&M management	23.1	24.0	25.7	27.2	10.7	52.5	30.6	6.2
Economic management	19.4	19.6	18.5	42.5	59.0	24.6	9.7	6.7
Environmental management	55.0	26.8	11.5	6.7	56.4	35.8	6.0	1.8
Mean	38.3	23.0	21.3	20.3	42.2	27.3	18.7	15.2
Max	64.7	41.5	66.9	42.5	79.5	52.5	46.3	80.2
Min	10.1	6.5	1.2	2.1	10.7	1.0	0.2	1.5
Water service level	44.0	36.9	16.1	3.0	15.1	63.1	20.2	1.6
Sanitation and hygiene service level	17.1	32.4	41.1	9.4	20.7	40.3	34.2	4.8
Water system infrastructure	35.9	39.1	22.4	2.6	31.3	51.6	14.7	2.4
Service provision	12.4	36.6	40.4	10.6	24.5	60.2	13.2	2.1
WaSH service level	14.8	51.9	31.2	2.1	12.1	59.7	27.3	0.9
Water services sustainability index	14.5	45.0	33.8	6.7	21.3	63.2	14.1	1.4
Wat. & San. service performance index	8.0	54.4	34.7	2.9	10.6	70.6	18.2	0.6
Mean	21.0	42.3	31.4	5.3	19.4	58.4	20.3	2.0
Max	44.0	54.4	41.1	10.6	31.3	70.6	34.2	4.8
Min	8.0	32.4	16.1	2.1	10.6	40.3	13.2	0.6

Table A4. Mean, median and interquartile range (IQR) values for the cases of Nicaragua (left) and Honduras (right).

NICARAGUA			
	mean	**median**	**IQR**
ACC	0.57	0.60	0.30
CON	0.70	1.00	0.67
SEA	0.79	1.00	0.33
QUA	0.68	1.00	1.00
SSL	0.55	0.61	0.62
PER	0.53	0.33	0.67
WAT	0.55	0.33	0.67
COM	0.44	0.33	0.33
AUT	0.43	0.38	0.71
INF	0.77	0.82	0.33
PRO	0.75	0.67	0.33
TRE	0.62	1.00	1.00
ORG	0.38	0.37	0.50
OPM	0.45	0.38	0.45
ECO	0.38	0.33	0.67
ENV	0.76	1.00	0.33

HONDURAS			
	mean	**median**	**IQR**
ACC	0.72	0.67	0.41
CON	0.75	0.83	0.00
SEA	0.81	1.00	0.33
QUA	0.17	0.00	0.00
SSL	0.63	0.70	0.41
PER	0.70	1.00	0.67
WAT	0.44	0.33	1.00
COM	0.51	0.50	0.33
AUT	0.68	0.79	0.54
INF	0.79	0.83	0.33
PRO	0.78	0.67	0.33
TRE	0.35	0.33	0.33
ORG	0.51	0.50	0.29
OPM	0.47	0.50	0.17
ECO	0.76	0.88	0.36
ENV	0.82	1.00	0.33

Figure A1. Primary networks resulting from the application of different structure learning algorithms and CIn tests. Top-left: fast.iamb (x2); top-right: fast.iamb (mi); bottom-left: inter.iamb (x2); bottom right: inter.iamb (mi). In black, input nodes. In blue, objective node. The final selection is highlighted.

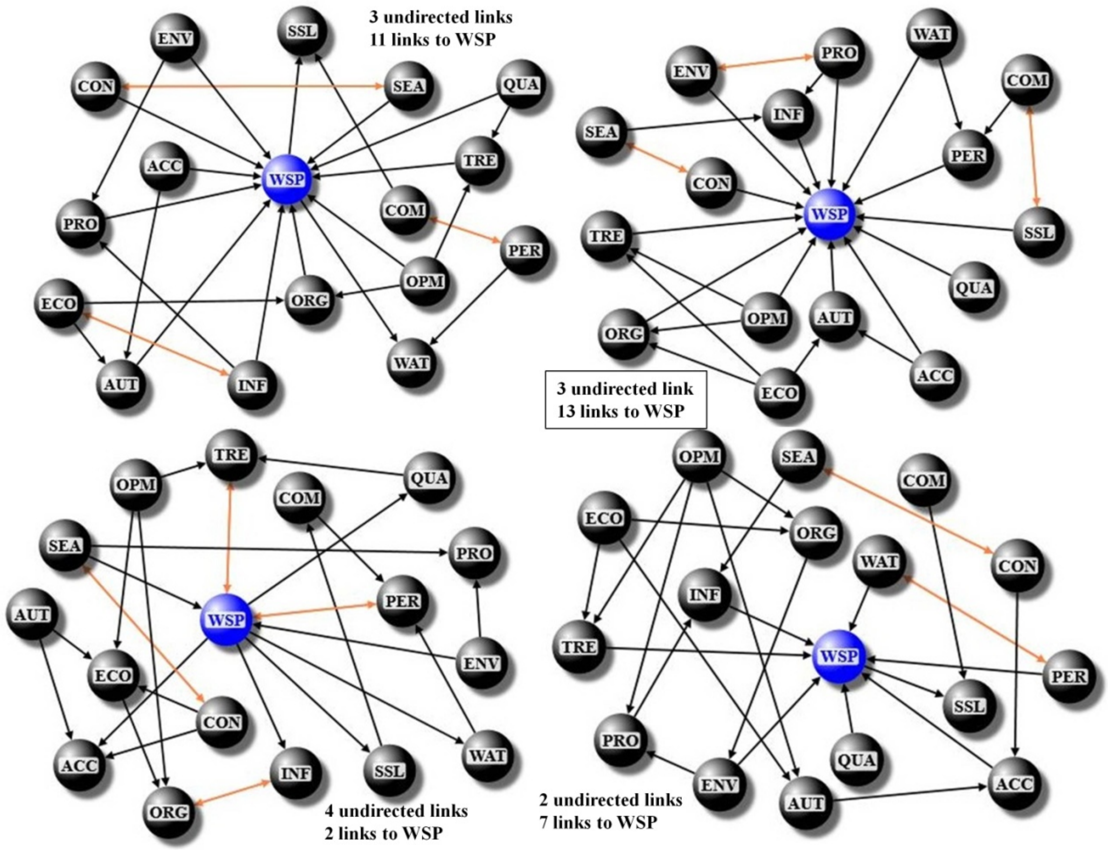
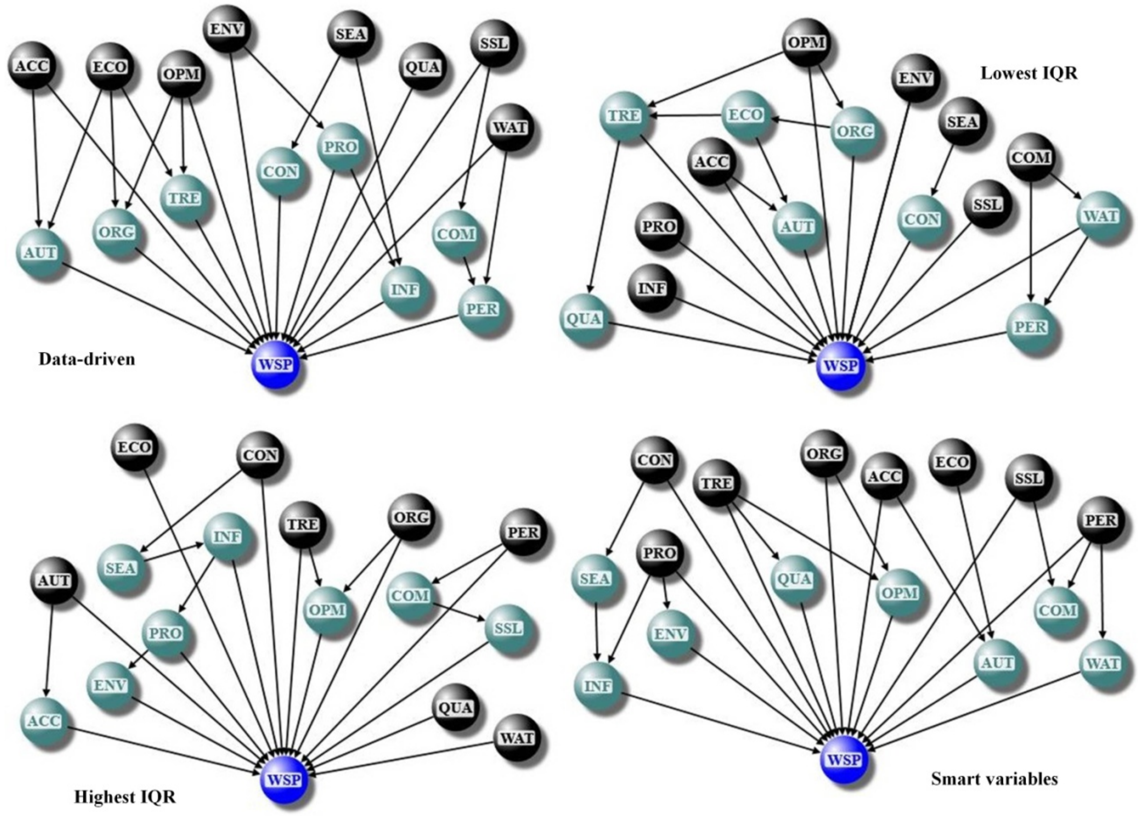


Figure A2. Candidate networks obtained from the tandem "fast.iamb" SLA and "mi" CIn test and assigning different inputs nodes according to the proposed approaches. In black, input nodes. In blue, WSP objective node.



.

Figure A4. . Full structure (expert knowledge) provision to networks after the iterative process proposed in the methodology. In black, input nodes. In red, nodes representing SIASAR conceptual model dimensions. In dark green, nodes representing SIASAR partial indices. In blue, WSP objective node. The final selection is highlighted (see Table S5 as well).

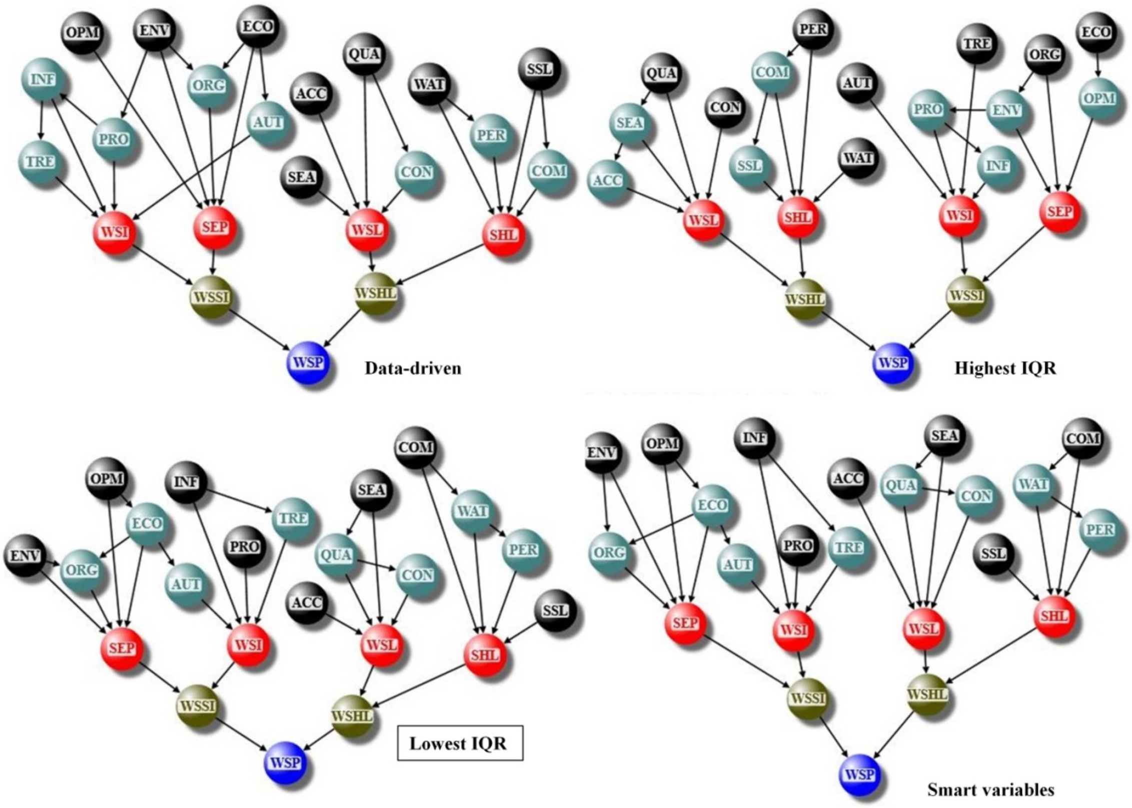


Table A5. Result bias against objective node "WSP" for the three scenarios considered and for the case of Nicaragua.

Scenario "16+1"					
Input nodes	**A (2.9%)**	**B (34.7)**	**C (54.4)**	**D (8%)**	**Method**
Data-driven	10.8	13.0	-21.4	-2.4	mle
	16.4	-29.5	-9.1	22.2	bayes
Lowest IQR	11.3	10.9	-20.1	-2.1	mle
	16.4	-29.4	-8.9	21.9	bayes
Highest IQR	12.0	11.2	-20.9	-2.3	mle
	16.5	29.6	-24.5	-21.6	bayes
Smart	11.0	15.3	-23.6	-2.7	mle
	16.4	-29.2	-9.1	21.9	bayes

All values are express in percentage.
"WSP" node is taking as a reference.

Scenario "16+2+1"					
Input nodes	**A (2.9%)**	**B (34.7)**	**C (54.4)**	**D (8%)**	**Method**
Data-driven	-1.2	14.2	-10.4	-2.6	mle
	0.8	-6.7	3.3	2.6	bayes
Lowest IQR	-1.2	12.6	-8.9	-2.5	mle
	0.9	-8.7	4.5	3.3	bayes
Highest IQR	-2.0	11.7	-7.3	-2.4	mle
	-0.3 (-2.7)	-3.4 (-9.3)	2.1 (17.0)	1.6 (-5.0)	bayes
Smart	-1.5	11.9	-8.0	-2.4	mle
	-0.2 (2.6)	-3.6 (-9.7)	2.1 (2.0)	1.7 (5.1)	bayes

All values are express in percentage.
"WSP" node is taking as a reference.
In brackets, biases obtained when using the test dataset for network validation.

Scenario "16+4+2+1"					
Input nodes	**A**	**B**	**C**	**D**	**Method**
Data-driven	-3.8	7.5	-1.7	-2.0	mle
	-3.7	6.4	-0.9	-1.8	bayes
Lowest IQR	-3.8	6.3	-0.6	-1.9	mle
	-3.7 (-2.5)	5.4 (2.8)	-0.2 (0.8)	-1.5 (-1.1)	bayes
Highest IQR	-4.3	7.4	-1.2	-1.9	mle
	-4.0	6.8	-1.1	-1.7	bayes
Smart	-4.0	6.9	-1.0	-1.9	mle
	-3.9 (-2.9)	5.9 (2.9)	-0.3 (1.1)	-1.7 (-1.1)	bayes

All values are express in percentage.
"WSP" node is taking as a reference.
In brackets, biases obtained when using the test dataset for network validation.

Table A6. Result bias applying Honduras database to Nicaragua final network. General index, partial indices and dimensions are considered for the comparison.

Nodes	A	B	C	D
Wat. & San. service performance index (WSP)	-3.6	-3.2	7.0*	-0.2
WaSH service level (WSHL)	6.6*	-0.7	-5.6*	-0.3
Water services sustainability index (WSSI)	-9.4	-7.1*	15.8*	0.7
Water service level (WSL)	36.0*	-27.0*	-9.1*	0.1
Sanitation and hygiene service level (SHL)	-0.1	-1.8	4.2	-2.3
Water system infrastructure (WSI)	5.5*	-8.1*	3.3	-0.7
Service provision (SEP)	-13.2*	-16.4*	28.8*	0.8

Values are express as percentages, taking as a reference the IS distributions of each node.
* Errors higher than 5% threshold established.

Table A7. Result bias in relation to "WSP" node for the case of Honduras.

Scenario "16+1"					
Input nodes	**A (10.6%)**	**B (70.6%)**	**C (18.2%)**	**D (0.6%)**	**Method**
Data-driven	-0.2	15.7	-14.7	-0.8	mle
	13.6	-42.9	6.2	23.1	bayes
Lowest IQR	-1.6	17.2	-14.8	-0.8	mle
	13.8	-43.4	6.4	23.2	bayes
Highest IQR	-1.5	16.8	-14.5	-0.8	mle
	13.6	-42.5	5.9	23.0	bayes
Smart	0.3	15.8	-15.3	-0.8	mle
	13.4	-42.7	6.3	23.0	bayes

All values are express in percentage.
"WSP" node is taking as a reference.

Scenario "16+2+1"					
Input nodes	**A (10.6%)**	**B (70.6%)**	**C (18.2%)**	**D (0.6%)**	**Method**
Data-driven	-2.2	11.2	-8.2	-0.8	mle
	-1.2	-9.9	8.2	2.9	bayes
Lowest IQR	-3.6	11.0	-6.6	-0.8	mle
	-2.2 (-0.9)	-3.2 (-6.5)	4.2 (4.5)	1.2 (2.9)	bayes
Highest IQR	-2.6	11.7	-8.3	-0.8	mle
	-1.5	-9.3	8.0	2.8	bayes
Smart	-2.9	11.7	-8.0	-0.8	mle
	-1.6	-6.2	5.8	2.0	bayes

All values are express in percentage.
"WSP" node is taking as a reference.
In brackets, biases obtained when using the test dataset for network validation.

Scenario "16+4+2+1"					
Input nodes	**A**	**B**	**C**	**D**	**Method**
Data-driven	-3.6	7.4	-3.1	-0.7	mle
	-3.5	6.1	-2.1	-0.5	bayes
Lowest IQR	-3.5	7.3	-3.1	-0.7	mle
	-3.6 (-3.4)	6.5 (5.9)	-2.3 (-2.6)	-0.6 (0.1)	bayes
Highest IQR	-3.7	7.6	-3.2	-0.7	mle
	-3.7 (-3.5)	6.7 (6.2)	-2.4 (-2.8)	-0.6 (0.1)	bayes
Smart	-3.4	7.2	-3.1	-0.7	mle
	-3.5 (-3.4)	6.1 (5.5)	-2.1 (-2.2)	-0.5 (0.1)	bayes

All values are express in percentage.
"WSP" node is taking as a reference.
In brackets, biases obtained when using the test dataset for network validation.