



Project Title:	ECOPOTENTIAL: IMPROVING FUTURE ECOSYSTEM BENEFITS THROUGH EARTH OBSERVATIONS
Project number:	641762
Project Acronym:	ECOPOTENTIAL
Proposal full title:	IMPROVING FUTURE ECOSYSTEM BENEFITS THROUGH EARTH OBSERVATIONS
Type:	Research and innovation actions
Work program topics addressed:	SC5-16-2014: “Making Earth Observation and Monitoring Data usable for ecosystem modelling and services”

Deliverable No: 7.3

Toolbox for decision support

(Ecosystem service mapping with Bayesian Networks – a manual)

Due date of deliverable:	30 th November 2018
Actual submission date:	11 th January 2019
Version:	v1
Main Authors:	ETHZ, UGR, UB, EBD-CSIC, Deltares



This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 641762



Project ref. number	641762
Project title	ECOPOTENTIAL: IMPROVING FUTURE ECOSYSTEM BENEFITS THROUGH EARTH OBSERVATIONS

Deliverable title	Toolbox for decision support (<i>Ecosystem service mapping with Bayesian Networks – a manual</i>)
Deliverable number	D7.3
Deliverable version	v1
Contractual date of delivery	30 th November 2018
Actual date of delivery	11 th January 2019
Document status	Final
Document version	V1
Online access	gbay.ethz.ch
Diffusion	Public
Nature of deliverable	Other
Workpackage	WP7
Partner responsible	Eidgenössische Technische Hochschule Zürich (ETH)
Author(s)	Stritih A., Celio E., Rabe S.-E., Robaina O., Moreno-Llorca R., Cazacu C., Adamescu C. M., Méndez P.F., Santamaría L., Wanke S., El Serafy G., Ziemba A., Grêt-Regamey A.
Editor	
EC Project Officer	Gaëlle Le Bouler

Abstract	This Deliverable presents a toolbox for ecosystem service modelling and mapping with Bayesian Networks (BNs). It contains guidelines on how to develop BN models of ecosystem services and an online platform (gBay) where users can run these models with spatial data. BNs can be used to map ecosystem services, understand the associated uncertainties and trade-offs, model future scenarios, or for participatory modelling with stakeholders, and we illustrate these different approaches with case studies from the Ecopotential project. We demonstrate methods to integrate different types of information,
-----------------	---



	including Earth Observation and in-situ data, stakeholder knowledge, and empirical models into BNs. In addition, users are provided with examples of BNs for ecosystem service mapping that can be downloaded and used for learning.
Keywords	Ecosystem Services, Bayesian Networks, Bayesian Belief Networks, Uncertainty, Mapping, Expert Knowledge, Stakeholders





Table of Contents

1.	Executive summary	7
2.	Introduction	9
3.	Developing a Bayesian Network	10
3.1	Model purpose and context.....	11
3.2	Bayesian Network software	12
3.3	Defining the nodes	12
3.3.1	With experts or stakeholders.....	12
3.3.2	Creating nodes in Netica	13
3.4	Designing the network	13
3.4.1	Defining links with stakeholders or experts.....	13
3.4.2	Creating links in Netica	14
3.5	Defining the variable states	14
3.5.1	Defining node states in Netica	14
3.6	Filling the CPTs	14
3.6.1	With experts or stakeholders.....	15
3.6.2	Linking remote sensing proxies to the state of the ecosystem	16
3.6.3	From existing empirical models	17
3.6.4	Learning from data or simulations.....	17
3.7	Testing, evaluating, and updating the BN	17
3.7.1	Sensitivity analysis	17
3.7.2	Scenario testing with experts or stakeholders	18
4.	Spatial BNs with gBay.....	19
4.1	Upload a Bayesian Network	19
4.2	Configure the run	19
4.2.1	Select target nodes	19
4.2.2	Set non-spatial evidence.....	19
4.3	Spatial inputs.....	19
4.3.1	Raster	19
4.3.2	Vector.....	20
4.4	Run the network	20
4.5	Outputs	21
4.5.1	Raster	21
4.5.2	Vector.....	22



4.6	Dynamic BNs	23
4.7	Accounting for spatial interactions	24
5.	Example Bayesian Networks	25
5.1	Carbon storage and timber production	25
5.2	Recreation	26
5.3	Habitat suitability	27
6.	Case studies	28
6.1	Avalanche protection in Davos, Switzerland.....	28
6.1.1	Developing the Bayesian Network.....	28
6.1.2	Quantifying the network (CPTs).....	29
6.1.2.1	Links between remote sensing proxies and the actual state of the ecosystem	29
6.1.2.2	Learning from process-based models.....	31
6.1.2.3	Incorporating empirical models.....	31
6.1.2.4	Expert knowledge: linking quantitative variables to qualitative categories.....	32
6.1.2.5	Expert knowledge: estimating distributions	33
6.1.3	Spatial application.....	34
6.1.4	Validation and sensitivity analysis	34
6.1.5	Description of the network nodes	36
6.2	Land-use scenarios in Sierra Nevada	39
6.2.1	Development of the Bayesian Network.....	39
6.2.2	Quantifying the network (CPTs).....	40
6.2.2.1	Stakeholder knowledge	41
6.2.2.2	Expert knowledge	42
6.2.3	Spatial application.....	42
6.2.4	Dynamics.....	43
6.2.5	Description of the network nodes	43
6.3	Collaborative modelling of ecosystem services in Doñana.....	46
6.3.1	Stakeholder engagement.....	46
6.3.2	Developing the Bayesian Network.....	47
6.3.3	Quantifying the network (CPTs).....	48
6.4	The impact of aquatic ecosystems provisioning services on tourism in the Danube Delta.....	50
6.4.1	Development of the Bayesian Network.....	50
6.4.2	Quantifying the network (CPTs).....	51
6.4.3	Spatial application.....	51
6.4.4	Description of the network nodes	52



6.5	A Trade-off Analysis of Ecosystem Services and Dynamic BNs in the Dutch Wadden Sea, the Netherlands.....	53
6.5.1	Developing the Bayesian Network.....	53
6.5.2	Quantifying the network (CPTs).....	54
6.5.2.1	Data analysis	54
6.5.2.2	Expert knowledge	55
6.5.3	Spatial application.....	55
6.5.4	Dynamics.....	55
6.5.5	Validation and sensitivity analysis	56
6.5.6	Description of the network nodes	56
6.5.6.1	Using Bayesian Networks to Capture Trade-offs Between Ecosystem Services.....	56
6.5.6.2	Using Dynamic Bayesian Networks to Capture Cumulative Effect on Ecosystem Services.....	60
6.6	Whale watching (Pelagos).....	64
6.6.1	Developing the Bayesian Network.....	64
6.6.2	Quantifying the networks (CPTs)	65
6.6.3	Spatial application.....	65
6.6.4	Validation and sensitivity analysis	65
6.6.5	Description of the network nodes	66
7.	References	68



1. Executive summary

Mapping and modelling ecosystem services (ESs) requires integrating different types of information about socio-ecological systems, including Earth Observation and in-situ data, empirical or process based models, and socio-economic data. In addition, local experts and stakeholders often have valuable knowledge about their socio-ecological systems, and involving them in the modelling process can facilitate communication and learning. Bayesian Networks (BNs) are a powerful tool for ecosystem service modelling, since they can integrate quantitative and qualitative data, including expert or stakeholder knowledge. In addition, the probabilistic structure of the BNs explicitly takes into account uncertainties, which are often high in ES assessments, and their graphic representation enables communication about the modelled system.

This toolbox aims to facilitate ecosystem service modelling with BNs by providing a tool to link BNs to spatial data, as well as guidelines and practical examples of how to develop, quantify, and validate BN models of ecosystem services.

In the first part, we introduce Bayesian Networks and provide guidelines for creating a BN through the following steps:

1. Defining the model purpose: BNs can be used to map current ESs, model future scenarios, analyse trade-offs and uncertainties, or develop a common understanding of the system with stakeholders. The next steps in the BN development will depend on the purpose and the question that we aim to address.
2. Defining the nodes: The most important variables in the system (such as key ecosystem services, their drivers, and Earth Observation proxies) can be defined through literature review or through a participatory process with experts or stakeholders.
3. Designing the network: The causal relationships between variables are described as directed links in the network, which can be developed based on literature, by experts or with stakeholders.
4. Defining the variable states: A set of possible states needs to be defined for each node in the network.
5. Quantifying the links: The links between nodes are quantified in conditional probability tables, which can be filled using data, models, or expert knowledge.
6. Testing and evaluating the BN: Once the network is compiled, it can be validated by testing different scenarios and observing the outcomes, by stakeholders or experts. It can also be analysed using sensitivity analysis.

For each step, we describe the different approaches that can be used, and illustrate them with examples from case studies developed in Ecopotential. We also show the practical implementation of each step in Netica, the BN software most commonly used among ecosystem service modellers, while providing information on other software options.

Next, we present gBay, an online platform where users can run their BNs with spatial data. In addition to performing inference using raster or vector geodata, gBay enables dynamic BNs – i.e. running the network over multiple time steps, where the output of one time step is an input to the next. Furthermore, advanced users can modify spatial data directly in gBay, which allows them to take into account spatial interactions (e.g. neighbourhood effects) that are often relevant in ecosystem service modelling.

To help users who are getting started with BN modelling, we provide examples of simple BNs that describe ecosystem services (carbon sequestration and storage, wood production, recreation, and habitats). The networks are based on real ecosystems in the Swiss Alps, but simplified. They can be downloaded, tested, and modified. In addition, we provide spatial data that can be used to test gBay with these networks.



Finally, the toolbox includes a set of case studies from the Ecopotential project, where BNs are being used to address a variety of questions related to ecosystem services. In Davos (Switzerland), Earth Observation, process-based and empirical models, and expert knowledge were combined in a BN of avalanche protection, and sensitivity analysis was used to identify the most important sources of uncertainty in the model. In the Sierra Nevada, stakeholder and expert knowledge is used to model future land use scenarios, while a collaborative modelling process in Doñana aims to build a better understanding of factors influencing the waterbird community. An expert-based BN of the Danube Delta explores trade-offs between provisioning and cultural services, and similar trade-offs are also addressed in the Wadden Sea by combining expert knowledge and in-situ data. Also for the Wadden Sea, a dynamic BN is used to assess the cumulative effect of pressures on mussel abundance over time. In the Pelagos Marine Protected area, a BN was developed with local experts to better understand the ecosystem service of whale-watching in the area, and its effects on the cetacean populations. For all the case studies, the networks and the methods used to develop them are described in detail, which will facilitate the application of similar approaches in other areas.



2. Introduction

Human well-being depends on a wide range of benefits and services provided by ecosystems, from carbon sequestration, protection from natural hazards, production of food and materials, to cultural services such as recreation and wildlife observation. The assessment of these ecosystem services (ES) and the trade-offs between them has been recognized as an important tool for conservation, ecosystem management, and spatial planning. However, modelling and mapping ES is often challenging, as it requires combining information on the state of the ecosystem, its functions, and the socio-economic demand for ES. For these different components, information may be available in different forms, including Earth Observation data, in-situ measurements, process-based models, and expert or stakeholder knowledge.

Bayesian networks (BNs) are a powerful tool to represent complex socio-ecological systems, as they can take into account both qualitative and quantitative data, while the associated uncertainties are explicitly shown and propagated through the network. Furthermore, relationships between variables in the BN are represented graphically, creating a transparent model structure that can facilitate communication with stakeholders. Because of these advantages, BNs have been used to support management decisions on watersheds (Ames et al. 2005, Bromley 2005), marine landscapes (Stelzenmüller et al. 2010), agriculture (Kleemann et al. 2017) and conservation (Newton et al. 2007). Other uses include environmental impact assessments (Marcot et al. 2001), habitat suitability modelling (Hamilton et al. 2015), risk assessment (Grêt-Regamey and Straub 2006), and land use change modelling (Celio et al. 2014). Increasingly, BNs are also being used to model ecosystem services (Pérez-Miñana 2016).

In this manual, we present guidelines and examples of how to model and map ES using Bayesian Networks. First, we describe the methods used to develop BNs, with practical examples. Then, we present an online platform (gbay.ethz.ch), where BNs can be linked with spatial data, and a set of example networks that can be used for learning. Finally, a set of case studies illustrate the use of BNs to address a variety of questions related to ecosystem services.

This Deliverable is closely linked to other work packages within the ECO POTENTIAL project. Spatial Bayesian Networks of ecosystem services can integrate Earth Observation products, such as those developed in WP4, process-based and empirical models (WP6; see Section 6.1), as well as future scenarios (WP8). By addressing trade-offs between ES and other questions relevant to management, and by involving stakeholders in the BN modelling process (e.g. in the Doñana PA, see Section 6.3), we can produce results that address the needs of PA managers and decision-makers (WPs 11 and 9).

The gBay platform for spatial BNs is currently password-protected and available only to ECO POTENTIAL partners, while in the current test phase. After publication (estimated in summer 2019), the platform will be openly available to all non-commercial users, and the underlying code will be published on GitHub. A link to the platform will then also be provided on the Virtual Laboratory Platform (WP10). The platform is accompanied by a Wiki (wiki.gbay.ethz.ch), where the contents of this Deliverable (instructions on how to develop BNs and how to use gBay, examples, and case studies) will be published, along with tutorials, example networks and data. Through the intuitive graphical user interface (GUI) of gBay and the material published on the Wiki, we aim to make spatial BNs an accessible tool, where users (including non-academic users, such as park managers – link to WP12) can learn how to use spatial BNs to address relevant questions in their socio-ecological systems.



3. Developing a Bayesian Network

A Bayesian Network is a directed, acyclic graph with an underlying joint probability distribution. It therefore consists of three parts:

1. A set of **nodes** representing variables, each with a set of mutually exclusive (discrete or continuous) states.
2. A set of **links** representing the (directed) causal relationships or dependencies between these nodes.
3. A set of **probabilities** specifying the belief that a node X will be in a particular state for each combination of its parent nodes' Pa(X) states (conditional probability tables - CPTs).
- 4.

Once the BN is compiled (completed, with all the CPTs filled), we “run” it by adding **evidence** to one or more nodes, which is propagated through the network and the joint probability distribution is updated. This calculation is called **inference**, and it is performed by applying Bayes' theorem:

$$P(X) = \sum_{Pa(X)} P(X|Pa(X)) * P(Pa(X)).$$

Evidence can be data (e.g. when we know the type of land cover in a pixel) or scenarios (e.g. when we explore what happens in the system when we increase the harvesting rate). When we know the state of a node with 100% certainty, this is called **hard evidence** (the land cover is a forest), while **soft evidence** contains some uncertainty and is in the form of a probability distribution (the land cover is a forest with 70% probability and a grassland with 30% probability).

In the following, we describe the procedure to develop a Bayesian Network to model ecosystem services.

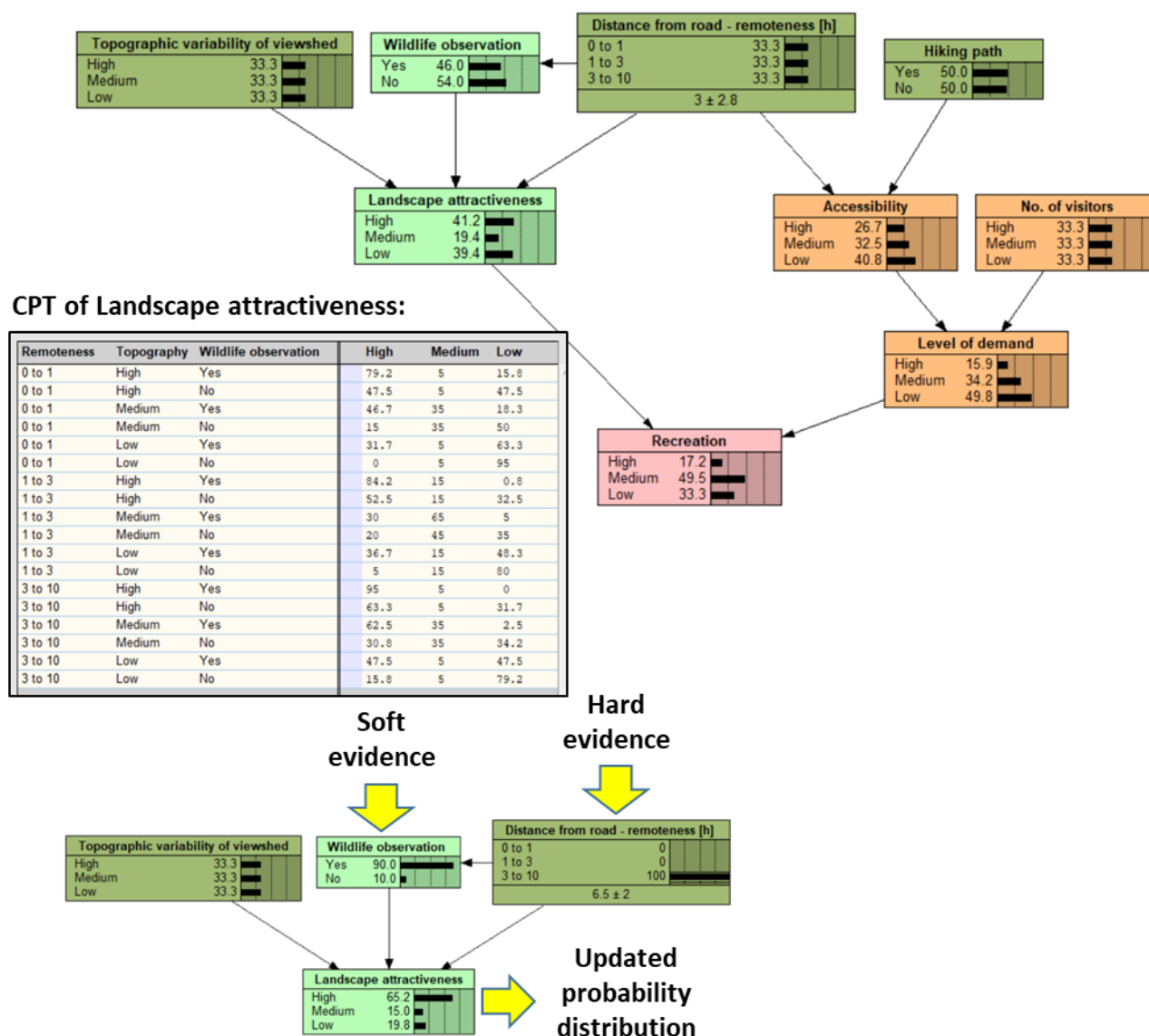


Figure 3.1-1: Example of a Bayesian Network that describes the ecosystem service of recreation. The CPT of node "Landscape attractiveness" is shown, as well as how the probability distribution of this node is updated when we add evidence to its parent nodes. For more details on this network, see [section 5.2](#).

3.1 Model purpose and context

Before the construction of the BN begins, it is important to clarify the problem and objectives, as well as the modelling context:

- What is the **main question** to be addressed with the model? What are the “target nodes” – the main variables of interest, e.g. the focal ecosystem services?
- What are the **key drivers** that affect the focal ecosystems?
- **How will the model be used?** BNs can be used to improve system understanding, for scenario or trade-off analysis, to support and inform management, to understand uncertainties, or for several of these purposes. The complexity of the model and the methods used to construct it should be adapted to its aim and target audience. For example, a simpler BN is more suited to communicate with stakeholders, while a more complex



model may help integrate various data and models, to produce more accurate ES maps. Furthermore, if a BN is aimed at communicating with stakeholders or managers, these should be involved in the development of the network and agree with its purpose (see example of [collaborative modelling in the Doñana PA](#)).

- What is the appropriate **spatial scale and extent** to address the question?
- Should the model be dynamic, i.e. take into account multiple **time steps**? If so, what is the appropriate temporal resolution?

3.2 Bayesian Network software

A variety of Bayesian Network software packages is available, such as Hugin, AgenaRisk, Genie, OpenBUGS, and various R packages. In this manual, we demonstrate how to create a BN in Netica, the BN software most commonly used in ecosystem service modelling. A free test version of Netica (with a limit of 15 nodes per network) can be downloaded at <http://www.norsys.com/>, and an extensive web help is available at: https://www.norsys.com/WebHelp/NETICA.htm#NETICA/X_Glossary_AE.htm

However, many functionalities are also available in other software packages. For an overview, see review by Pérez-Miñana (2016).

In order to map ES, we link BNs to spatial data using a specialized online application (**gBay.ethz.ch**) developed at ETH-PLUS within the framework of the ECOPOTENTIAL project.

3.3 Defining the nodes

Once the purpose and context of the BN are defined, we select the variables that should be represented as nodes in the network in order to address the question. The BN can include the following sets of variables:

- Focal ecosystem services as target nodes
- Ecosystem properties (structure and processes) that influence the provision of these services
- Controlling factors that affect the relevant ecosystem properties (drivers, pressures)
- Variables describing the demand for the ecosystem service
- Management interventions and scenarios
- Proxies and indicators for variables that cannot be directly assessed, including remote sensing variables

The relevant variables in the system can be defined by experts, from literature, or through stakeholder workshops.


3.3.1 With experts or stakeholders

A good way to find a set of relevant influence factors is a workshop with experts or stakeholders. Before the workshop, a list of relevant factors can be compiled based on literature or expert interviews. At the workshop, the participants individually rank the importance of the individual factors. Then, the ranking is discussed to find a consensus about a set of most relevant influence factors (Celio et al. 2014). For more details on setting up a BN with experts or stakeholders, see Cain (2001), Celio et al. (2014), and Salliou et al. (2017).

Workshops with stakeholders and experts were used to develop a BN model of land use change in the [Sierra Nevada](#), and to model ecosystem services in [Doñana](#).



3.3.2 Creating nodes in Netica

To create a node in Netica, click on the  icon and then place the new node somewhere on the canvas. Double click on the node to open a window with the node properties, where you can change the node name and define it as discrete or continuous.

3.4 Designing the network

In the next step, the relationships between the nodes are included as **directed links** in the network. The links represent causal relationships, not to be mistaken with the flow of information. For example, an EO product such as a land cover classification is an input node to the network, as it provides information about the actual land cover. However, in terms of causality, the classification depends on the actual state, not vice versa, so the link should be directed from the land cover to the classification. Defining links based on causality maintains the logical structure of the network. However, it is not always easy to define causality, and sometimes a pragmatic approach is necessary, where links are defined in a way that facilitates the construction of CPTs.

During this step, we follow the following **guidelines**:


- Nodes with no parents should represent the state of the ecosystem, driving factors, or management interventions.
- Nodes with no children should be the ES of interest or additional impacts (these only if their inclusion is relevant to the modelled problem).
- If possible, there should not be more than three parents for each child node (the size of a CPT grows exponentially with the number of parents).
- **Feedback loops** are not allowed in BNs. In case of a feedback effect, consider which direction of impact is more relevant for the modelling purpose and within the temporal and spatial scale of interest. The alternative is to use a dynamic BN, consisting of several BNs at different time steps (see section on dynamic BNs).

The resulting network should be **reviewed** by other experts, and (if aimed for communication and management) by the relevant stakeholders. During the process, it is often necessary to return to the previous step and define new nodes or redefine existing nodes.

3.4.1 Defining links with stakeholders or experts

One approach to define the causal links between nodes is using an influence matrix, where experts (or stakeholders) are presented with a matrix based on the list of nodes that they previously agreed on. Then, they are asked to assess the influence of each node on every other node in the network (Celio et al. 2014). This can be done in a workshop or via an (online) questionnaire. The influence matrix can be translated into causal links in the network, which should then be discussed with the experts. As a first step to construct a network, it is necessary, to construct the casualties of all driving factors. Because there is no natural scale for judging direct impact strength, the scale has to be determined, e.g. a 5-point-scale. A problem that might appear is the high amount of impacts. To deal with this phenomenon, the calculation of standard statistics (mean, median, variance) are helpful to show the degree of heterogeneity and to aggregate the experts impact estimations. Based on this a preliminary network can be constructed to be discussed with the experts.

3.4.2 Creating links in Netica

To create a link between two nodes in Netica, click on the  icon in the toolbar. Then click on the parent node and drag the arrow to its child node.

3.5 Defining the variable states

For every node in the network, we define whether they should be described as **discrete** (states are categories, such as land cover) or **continuous** (e.g. biomass) variables. For discrete variables, a finite set of mutually exclusive states should be defined. The number of states exponentially increases the complexity of the CPTs, and should therefore be kept to a minimum that still represents the states relevant for the system. Continuous variables need to be discretized into intervals, where the number of intervals should be as low as possible, while accounting for any important thresholds and maintaining the shape of the probability distribution.

3.5.1 Defining node states in Netica

To modify the states of a node in Netica, double click on the node to open the node properties window. Click on “Description” and select “States”. Then, simply enter the names of the states in the box below. For continuous nodes, the states are defined through discretization intervals. Select “Discretization” and enter the threshold values between the states.

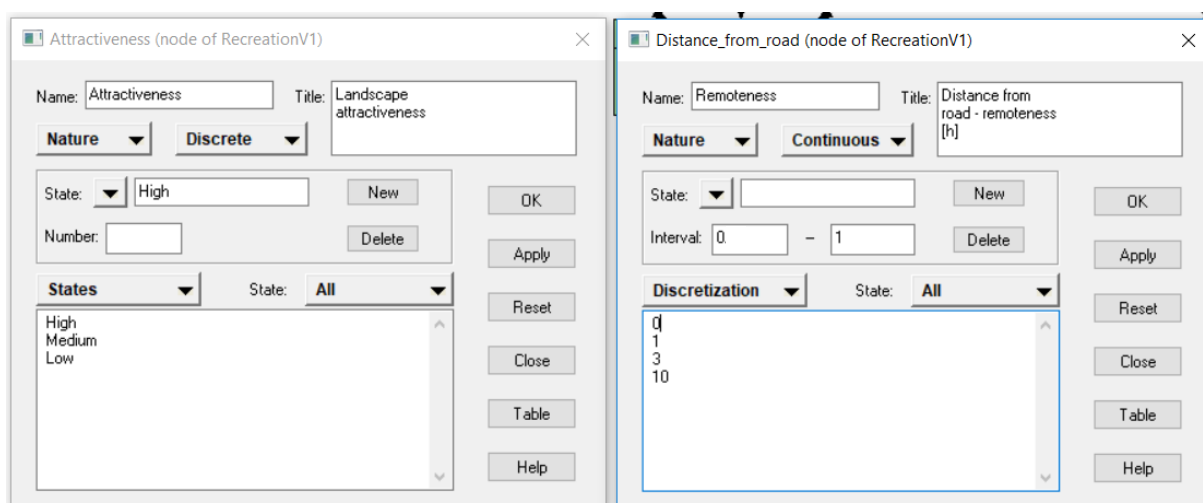



Figure 3.5-1: Defining states for discrete and continuous nodes in Netica.

3.6 Filling the CPTs

The links between nodes in a BN are represented by conditional probability tables, where a probability distribution of a child node is defined for every combination of states of its parent nodes. Depending on the availability of data or models, various methods can be used to populate CPTs, from expert elicitation to “learning” from data. During this process, we may find it necessary to return to previous steps and redefine the nodes, their states, or the links between them.

In Netica, the CPTs can be filled either **manually** (right click on the node -> Table) or through **equations**. To use an equation, open the node properties, select Equation, and enter the equation in the box. Then, click on  to calculate the CPT based on the equation.

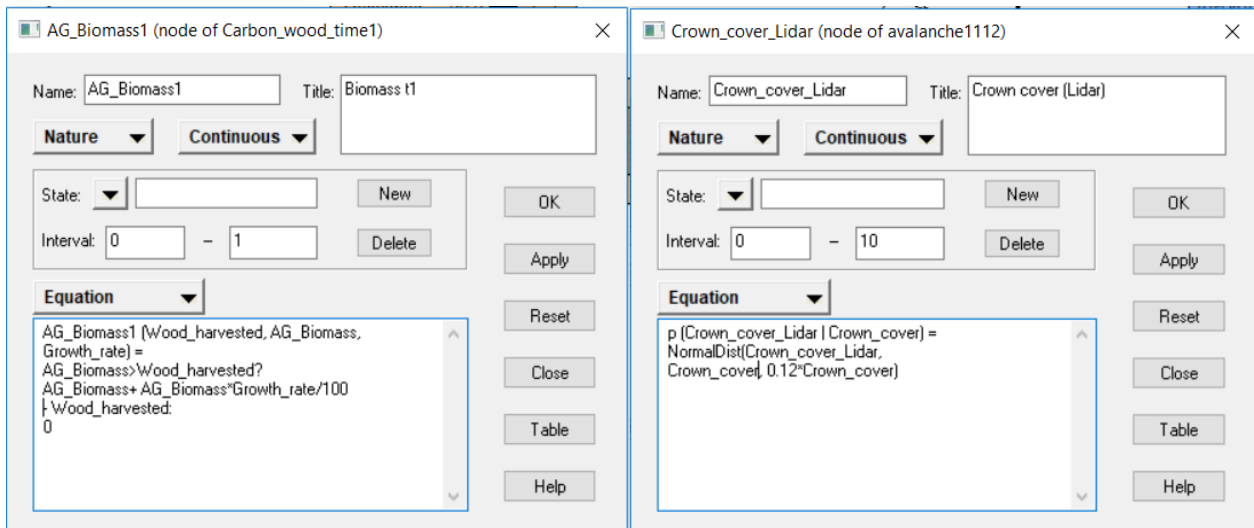


Figure 3.6-1: Examples of a deterministic (left) and probabilistic (right) equation in Netica.

3.6.1 With experts or stakeholders

When data is lacking, the CPTs can be filled manually by experts or by stakeholders. Usually, this means that the experts should specify the probability of each state of the child node given each combination of parent nodes. When a node has many parents with several states, many rows of CPTs need to be filled, which can lead to fatigue and boredom, and it is difficult to ensure consistent distributions (Das 2004). This is why it is important to limit the number of parents, and the number of node states.

When node states are binary or ordered, this problem can be reduced by using various interpolation methods (Cain 2001, Das 2004). For example, if a child node C has three states (“low”, “medium”, “high”), and three parents Pa with the same three states, we can elicit three probability distributions of C: where all three parents are in state “high”, all in state “medium”, and all in state “low”. In addition, we need the relative weights of each parents (w_1 , w_2 , w_3 , which should sum up to one). Then, all the other rows of the CPT can be calculated by interpolation.

Table 3.6-1: Example of interpolation in a CPT, where the weights of the parents are $w_1 = 0.17$, $w_2 = 0.5$, $w_3 = 0.33$. Elicited probabilities are shown in bold, while all other probabilities in the table are calculated using interpolation, as shown below.

Parent nodes			Child node probabilities [%]		
Parent 1	Parent 2	Parent 3	High	Medium	Low
High	High	High	95	5	0
High	High	Med	= $95*w_1 + 95*w_2 + 10*w_3 = 66.7$	= $5*w_1 + 5*w_2 + 80*w_3 = 30$	= $0*w_1 + 0*w_2 + 10*w_3 = 3.3$
...
Med	Med	Med	10	80	10
...
Low	Med	Low	= $0*w_1 + 10*w_2 + 0*w_3 = 5$	= $5*w_1 + 80*w_2 + 5*w_3 = 42.5$	= $95*w_1 + 10*w_2 + 95*w_3 = 52.5$
Low	Low	Low	0	5	95

Distributions of continuous variables can also be elicited from experts. One useful approach is the four-point estimation method (Speirs-Bridge et al. 2010), where we ask experts for the expected value of the node for a specific combination of parents, the expected upper and lower bounds of possible values, and their confidence in their estimate. Using this information, we can estimate a probability distribution (e.g. a normal or triangular distribution). An example of this approach is available in the [avalanche protection case study](#).

In some cases, experts find it easier to deal with categories rather than continuous variables, and it may be useful to translate continuous nodes to discrete classes using [fuzzy logic](#).

When using expert elicitation, more than one expert should be involved whenever possible. Then, we can evaluate the differences between the experts, and test the sensitivity of the network to these differences. This may be particularly interesting when developing a network with different stakeholder groups. For a final network, the CPTs are usually combined by averaging the values from different experts.

Some stakeholders may find it difficult to estimate probabilities. In this case, surveys should be designed in a way that is understandable and easy to fill out. In order to understand farmers' decisions in the [Sierra Nevada, a choice experiment](#) was used.

3.6.2 Linking remote sensing proxies to the state of the ecosystem

To map ecosystem services, proxies of ecosystem structure are often derived from remote sensing (e.g. land cover classifications or LiDAR-based measurements of vegetation cover). However, these remote sensing products often include some uncertainty due to measurement errors or misclassifications. To make these uncertainties explicit, we can create separate nodes representing the observed value and the actual state of the variable. The observation is caused by the actual state, not vice-versa, and defining the structure of the network based on this causality helps to define conditional probabilities (see [avalanche protection case study](#)).



3.6.3 From existing empirical models

Often, some parts of the network have already been extensively researched and empirical or process-based models are available in literature. In this case, the model can be incorporated in the BN in the form of probabilistic equations. This usually means that the probability distribution of the child node is a normal distribution, where the mean is a function of its parents, and the standard deviation is derived from the reported uncertainty in the model. Other types of distributions can also be used.

For an example of how an empirical model can be incorporated in a BN, see the [avalanche protection case study](#).

3.6.4 Learning from data or simulations

Where sufficient data is available, CPTs can be “learned” directly from data within a BN software. Learning a CPT requires a dataset of cases with information on the child node and its parents. Various algorithms (e.g. Expectation Maximisation or gradient descent) can be used to find the maximum likelihood BN, which is the network that is most likely given the data. This approach can also be used to translate process-based simulation models into the BN. We run the simulation over the range of input values, and use the results as cases for learning the BN. Parameters that are not included in the network should also be varied in the simulation, in order to capture the uncertainty in the model.

To “learn” from data in Netica, create a text file where the column names match the names of the nodes in the network, and rows represent cases (i.e. observations, plots, measurements). Go to **Cases -> Learn -> Incorp Case File** (to derive CPTs by simply counting the cases) or **Learn Using EM** (to use the expectation maximisation algorithm, e.g. in case of missing data). To use learning only for the CPT of one node, select the node before performing the learning.

Learning from simulations was used to populate one of the nodes in the [avalanche protection network](#), while in-situ data were used to quantify some nodes in the BN of [ecosystem services in the Wadden Sea](#).

3.7 Testing, evaluating, and updating the BN

Once all the CPTs in the BN are filled, the network can be compiled. Then, it can be tested by entering “evidence” (modifying the probability of a node, e.g. setting the probability of a specific state to 100%). The evidence is propagated through the network by inference, and we can observe the resulting probabilities in other nodes. The BN and its behaviour under different scenarios should be presented to experts and stakeholders, and evaluated by them. Afterwards, their feedback is used to adjust the structure of the network, the states of the variables, or to update the CPTs.

3.7.1 Sensitivity analysis

Sensitivity analysis is a useful tool that determines the influence of individual variables on the outputs, which can help us decide where more data is necessary. When new data becomes available, it can be used to update the conditional probability distributions in an iterative process.

In Bayesian Network modelling, sensitivity analysis is often used to evaluate the influence of variables in the modelled system on the posterior probability distribution of a node of interest (Uusitalo 2007, Marcot 2012). Sensitivity to findings can be measured by the reduction in uncertainty (e.g. entropy or variance) in the target node



due to a finding on another node. Entropy reduction is expressed by the measure of mutual information (Kjaerulff and Madsen 2013):

$$I(X, Y) = H(X) - H(X|Y) = H(Y) - H(Y|X) = \sum_Y P(Y) \sum_X P(X|Y) \frac{P(X, Y)}{P(X)P(Y)}$$

where $H(X)$ is the entropy of X and $H(X|Y)$ is the entropy of X after a new finding on Y .

The analysis of sensitivity to findings gives us an indication of which variables in the system have the highest influence on the outcome of the model.

To perform sensitivity analysis in Netica, select the node whose sensitivity you want to measure. Then, go to **Network -> Sensitivity to findings**, and you will get a table showing the sensitivity metrics (mutual information, relative mutual information, variance of beliefs, and, for continuous nodes, variance reduction).

Sensitivity analysis can be used to map the flow of information in a Bayesian Network and analyse uncertainties (see [Section 6.1.4](#)).

3.7.2 Scenario testing with experts or stakeholders

When co-constructing a BN with experts or stakeholders, it is important that they are presented with the results of the process, and can validate them. In a workshop setting, the participants can input scenarios in the network, see the results, and evaluate whether these agree with their understanding of the system (Celio et al. 2014).



4. Spatial BNs with gBay

Mapping ecosystem services requires running the BNs in a spatially explicit way. Within ECO-POTENTIAL, we have developed a toolbox that links BNs to **spatial (raster or vector)** data. For each pixel (or polygon), the values in the input data are used as evidence in the network, and inference is performed to obtain the **posterior probability distribution of the target nodes**. Then, the posterior distributions of the target nodes are written into a new spatial file.

4.1 Upload a Bayesian Network

In the first step, the BN we want to run should be uploaded in the **.dne format** (as created in Netica or another BN software). This is done by using “choose file” and selecting the network file, or dragging and dropping the network file into the application. Then, we click “Proceed”.

Once a network has been uploaded, it is visualized in the GUI, displaying the nodes with their states (in case of continuous nodes, the upper bound of each interval is displayed), and the links between them. The nodes can be moved around.

4.2 Configure the run

4.2.1 Select target nodes

When hovering with your mouse above a node, additional options appear. A node can be selected as a **target node**, which means that the output of running the network will contain the posterior probability distribution of this node. When we select a node as a target node, a target icon will appear in its upper right corner.

You may want to save the **configuration** of the run (including the information about which are the target nodes, etc.), in case you will run it again later. You can also upload a previously saved configuration.

4.2.2 Set non-spatial evidence

By clicking on the state of a node, we set that state as hard evidence, which is then shown in bold. This is useful for **non-spatial evidences**, for nodes which have the same state over the whole study area, such as an agricultural policy.

4.3 Spatial inputs

4.3.1 Raster

In the raster mode of gBay, the input data as well as outputs are in raster (**.tif**) format. The input rasters correspond to nodes that we have data on. We can upload input rasters by simply dragging the **.tif** file from our file explorer to the corresponding node in gBay. When data is added to a node, the colour of the node changes, and the name of the **.tif** file is displayed.

The input rasters should all have the same extent, cell size, and spatial reference (coordinate system).



When the input data represents **hard evidence** (we know the state of the node at each pixel, with 100% certainty, e.g. we know that the land cover of a pixel is forest), then the input raster has one band, where the value of each pixel corresponds to the number of a state of the node. For continuous nodes, the value of the pixel represents the real value (e.g. forest cover of 75%).

When we use **soft evidence** (a probability distribution for each pixel), the input raster should have as many bands as the number of states of the corresponding node. Each band represents the probability of a state (e.g. the probability that the land cover of a pixel is forest). The values of all bands should sum up to 100.

4.3.2 Vector

When using vector data (.shp files or ESRI geodatabase files), the input nodes are represented in the attribute table of the dataset. For **hard evidence**, the **attribute table** should have a column corresponding to the name of the input node, with values corresponding to the states of the node.

For **soft evidence**, each state of the input node should be represented by a column of the attribute table, with values representing the probabilities of the states (which should sum up to 100). The column names should be the node name, followed by two underscores, an 's' (from state) and the number of the state. For example, to set soft evidence on node 'lu_t0', with three states, the attribute table should contain the columns: lu_t0__s1, lu_t0__s2, lu_t0__s3

The vector file can be added to the network by dragging it to the box labelled “**Vector file**”.

Please note that gBay does not modify the geometry of the vector file, but simply performs inference on each object, using information from the attribute table.

4.4 Run the network

Once you have set up the network, selected the target nodes, and uploaded the spatial inputs, you can click on “**Run**” to run the network. gBay will use your spatial data to perform inference in the BN for each pixel or object, and produce an output of the results.

Any potential **errors or warnings** will appear as pop-ups. If you want to see the progress of the processing, select the option to “Show console”.

If you have a complex network and are running it with a **large** spatial dataset, this may take some time. In this case, you should enter your email address, where you will receive a **notification** when the process is completed, along with a link where you can download the data.

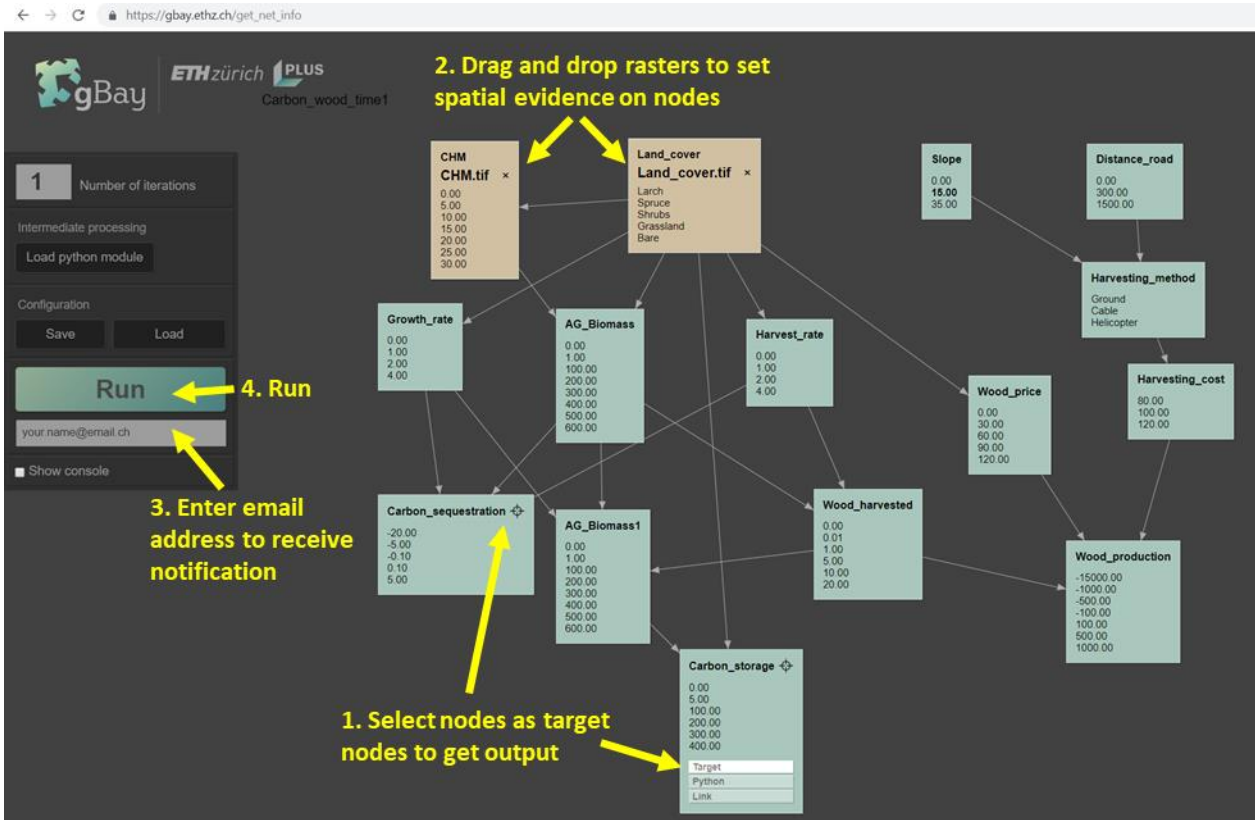


Figure 4.4-1: The gBay interface set up to run a BN to map carbon sequestration and storage (for details on the network, see [Section 5.1](#))

4.5 Outputs

4.5.1 Raster

The output of a gBay run is a posterior probability distribution raster of each target node (named target_node.tif). This raster has one band for each state of the target node, where the value represents the probability of the state. In addition, the last band of the raster represents the most probable state.

In addition, for each target node, some metrics of the posterior probability are calculated and stored in an additional raster file called target_node_stats.tif. For **discrete** target nodes, this file contains one band with **Shannon's evenness index** of the posterior probability distribution:

$J = H/H_{max}$, where $H = \sum_{i=1}^N p_i \cdot \log_2 p_i$, $H_{max} = \log_2(N)$, p_i is the probability of state i and N is the number of states. The index is a standardized measure of entropy (can be compared between nodes with different numbers of states) and expresses uncertainty. It has values between 0 and 1, where 1 denotes a uniform distribution between all possible states (maximum uncertainty), and 0 denotes complete certainty that the output node is in a specific state.

If the target node is **continuous**, the stats output contains six bands with the following values:

1. Evenness index
2. Mean
3. Median
4. Standard deviation

4.5.2 Vector

In vector mode, the gBay output is a vector file, where the geometry corresponds to the input geometry, and the attribute table contains the posterior probability distribution of the target nodes, with a column for each state, and an additional column with the most probable state.

Table 4.5-1: Overview of spatial inputs and outputs in gBay

	Input format	Input values	Output
Raster	.tif file per input node	Hard evidence: Value = node state (discrete nodes) or continuous value (continuous nodes)	target.tif: one band per state: value = probability of state additional last band: value = most likely state target_stats.tif: bands: 1. Evenness index Only for continuous nodes: 2. Mean 3. Median 4. Standard deviation
		Soft evidence: One band per state value = probability of state	
Vector	one .shp file or geodatabase .gdb (reads the attribute table)	Hard evidence: Column of attribute table with same name as input node value = node state (discrete nodes) or continuous value (continuous nodes)	Same geometry as input attribute table with a column for each state of the target node: value = probability of state additional column: value = most likely state
		Soft evidence: Columns with node name and state number (lu_t0__s1, lu_t0__s2, lu_t0__s3): value = probability of state	

4.6 Dynamic BNs

Bayesian Networks usually represent a static state of the studied system, and **feedback loops** cannot be incorporated. To take into account changes over time, including feedbacks, we can use dynamic BNs, using the so-called “time-slicing” approach. In practice, this means that each time step is represented by a run of the network, where the **outputs of one step are used as inputs to the next** time step.

For example, when modelling carbon storage over time, we start with an estimate of the forest biomass at the beginning of the first time step (Biomass). During one time step, growth and harvesting take place, and through inference we obtain the probability distribution of land use after the first time step (Biomass1, e.g. after 10 years). This Biomass1 then becomes the input for Biomass in the second time step.

In gBay, we create such a dynamic BN by specifying links across time steps. To do this, click the option “Link” on the output node that should feed into the input of the next time step (Biomass1). An orange arrow will appear, which should be connected to the corresponding input node (Biomass).

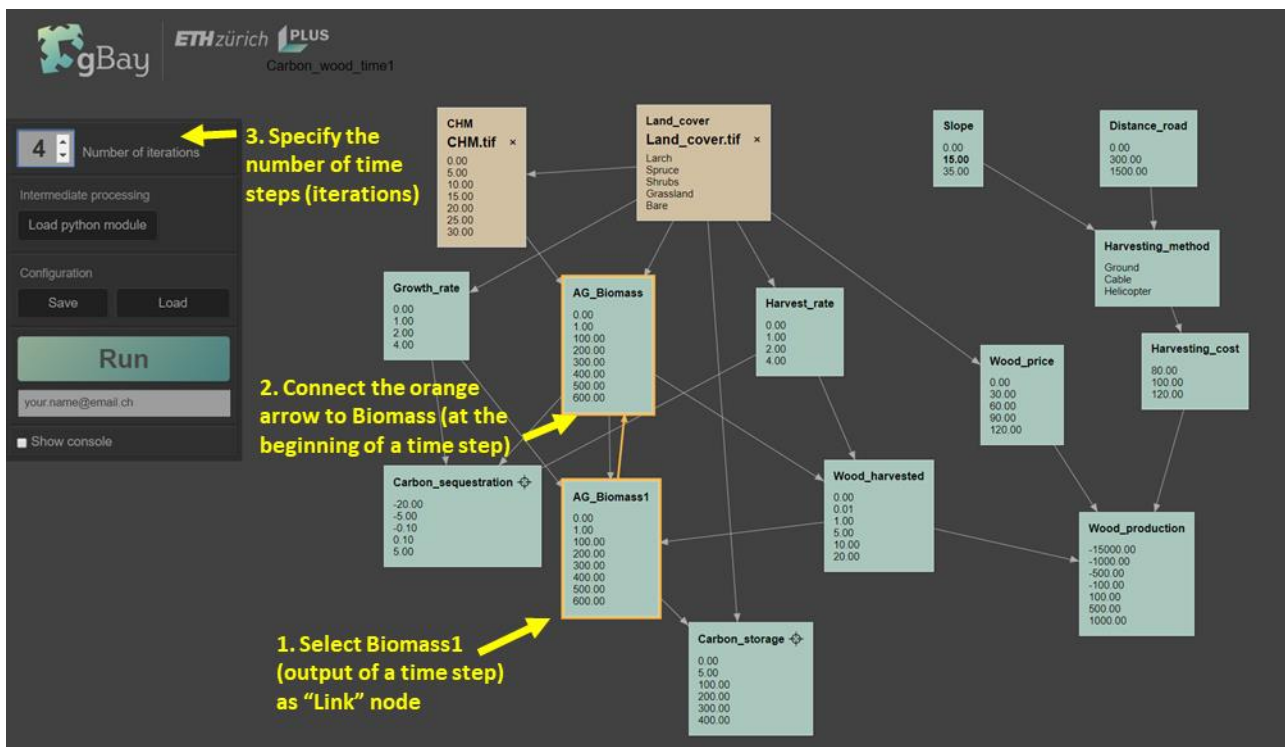


Figure 4.6-1: Using gBay to run a BN dynamically (over multiple time steps). For details on the network, see [Section 5.1](#)

Dynamic BNs are used to model scenarios of land use change in [the Sierra Nevada](#), and to understand the cumulative effects of pressures on mussels in [the Wadden Sea](#).



4.7 Accounting for spatial interactions

Spatial interactions, such as neighbourhood effects, are often important in socio-ecological systems. In gBay, we can account for spatial interactions using the **Intermediate processing** module. Here, we can upload **python scripts** that process available data on the nodes (either input data or outputs of the previous time step) and modify evidences according to some rules. Some examples are provided, including:

- **Changing evidences** over time (e.g. modifying the evidence on a node describing agricultural policy, which changes between different time steps).
- Implementing **boundary conditions** (e.g. limiting the number of pixels that can change from “forest” to “agriculture” under a specific land use change scenario).
- Calculating **neighbourhood** values (e.g. calculating the percentage of forest cover in neighbouring pixels within a specified distance).

To use the intermediate processing module, modify one of the available scripts with your parameters (such as neighbourhood distance) or upload your own. In addition, specify which nodes' data should be used for the calculation by hovering over the node and selecting the option “Python”.

5. Example Bayesian Networks

In the following, we present three examples of Bayesian Networks for modelling ecosystem services. The networks and some associated data can be downloaded (at <https://wiki.gbay.ethz.ch/doku.php?id=examples>) and used to get familiar with Netica and gBay. They are based on real ecosystems in the Swiss Alps, but are simplified, and need to be adapted with own data and/or knowledge in order to apply them elsewhere.

5.1 Carbon storage and timber production

In order to model carbon sequestration, storage, and wood production in mountain forests, we use two remote sensing products, a land cover classification and a LiDAR-based canopy height model (CHM). The CHM is linked closely related to the total biomass of a forest stand. From one time step to the next, the forest biomass changes due to growth and harvesting, and this change accounts for most of the net carbon sequestration in mountain ecosystems. Forest inventory data (*learning_biomass.txt*), which contains information about tree height, stand biomass, growth and harvest rates per plot, is used to “learn” the CPTs of these nodes.

To calculate total carbon storage, probability distributions of soil carbon for the different land cover types from literature are added to the aboveground biomass carbon.

The demand for wood production depends on wood prices, as well as harvesting costs, which are related to topography (slope) and accessibility (distance to roads).

This model can be run dynamically, over multiple time steps. The biomass at the end of one time step (Biomass t1) can be used as an input on Biomass t0 in the next time step.

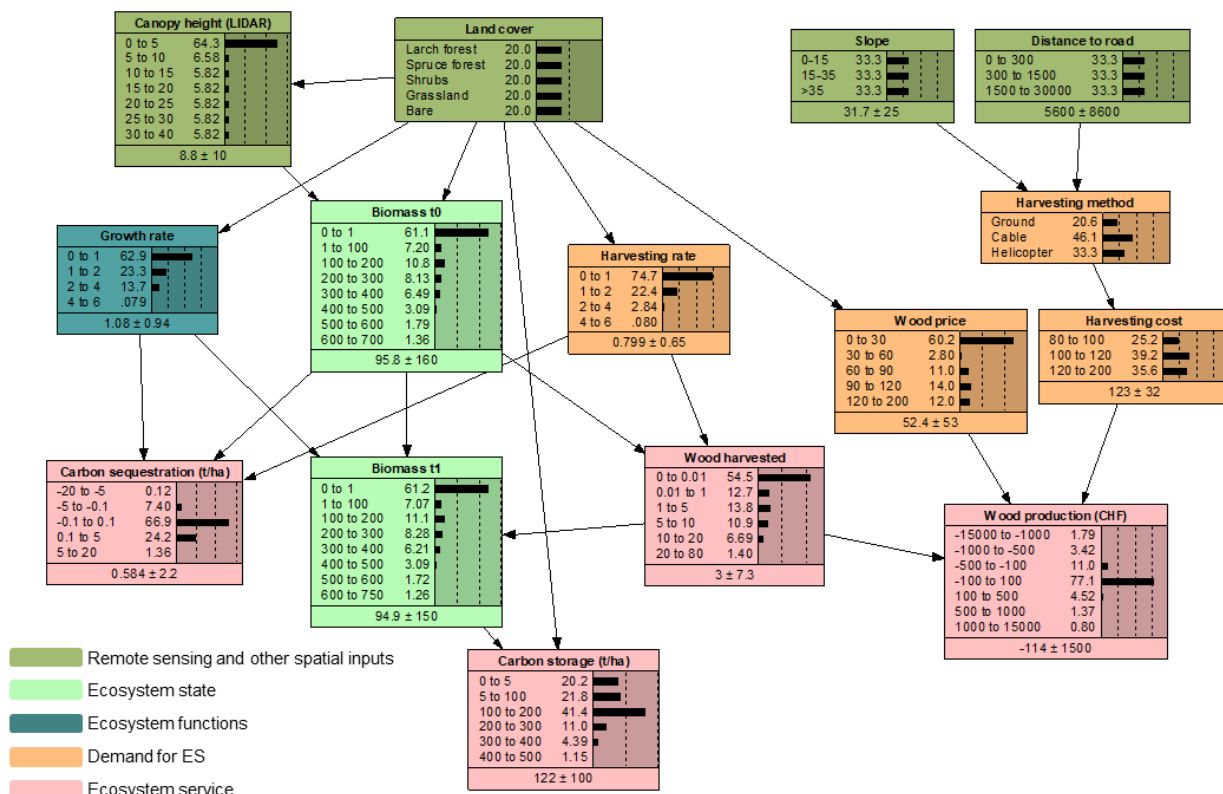


Figure 5.1-1: A Bayesian Network that models carbon storage, sequestration, and wood production in a mountain region

5.2 Recreation

The value of recreation in a landscape depends on its natural potential (i.e., the attractiveness of the landscape), as well as its accessibility and the overall level of demand (e.g. the number of visitors in the region). In this network, we combine three factors that were defined by local stakeholders in the mountain region as important for landscape attractiveness: the topographic variability (heterogeneous terrain with prominent peaks), remoteness, and probability of observing wildlife.

On the other side, the demand for recreation occurs in accessible areas (on hiking paths or near roads) and is higher when the number of visitors is high.

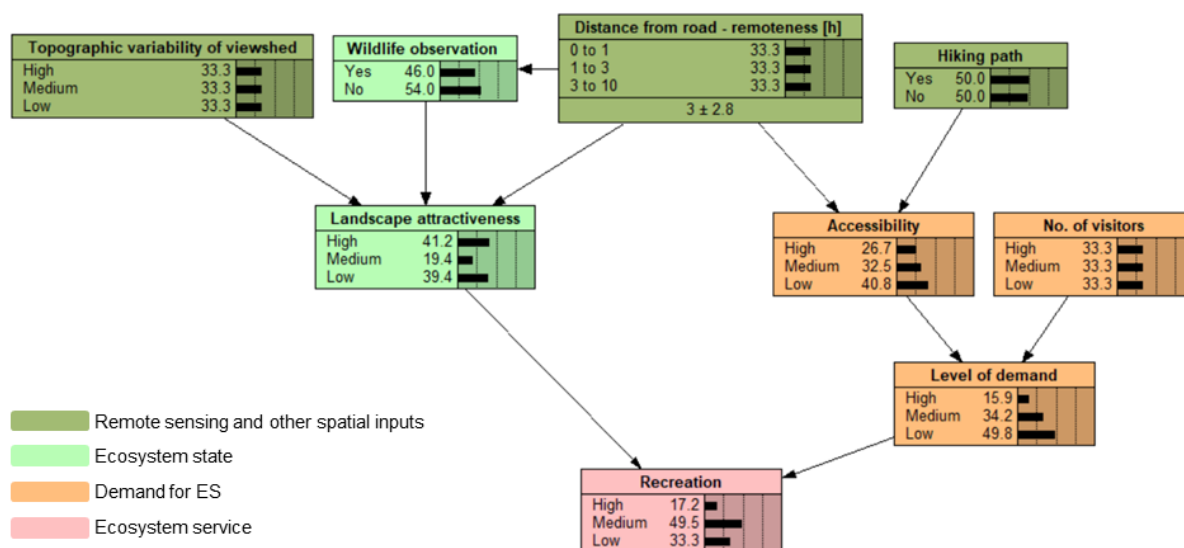


Figure 5.2-1: Bayesian Network describing the recreation ecosystem service in a mountain region.

5.3 Habitat suitability

The model describes the habitat suitability for a grouse species in the Alps. The grouse mainly live in montane, not-too-dense conifer forests (with 40-70% tree cover), where there is enough light for understory vegetation such as blueberries. Large forest areas are more suitable than small patches, so the model takes into account nearby pixels within a 500 m distance. Since the birds are sensitive to human disturbance, the distance from infrastructure is also an important factor.

For each continuous factor, we first define the thresholds to discretize the nodes. Then, we create additional nodes describing the suitability of each individual factor. Since all the factors need to be suitable in order to have a suitable habitat for grouse, the conditional probability of “habitat suitable” is defined as a Noisy-AND distribution of all its parents. In Netica, we write this using an equation:

$$p(\text{Habitat} \mid \text{Forest_cover_suit}, \text{Forest_cover_500suit}, \text{Distance_suit}, \text{Elevation_suit}) = \text{NoisyAndDist}(\text{Habitat}, p_{\text{inh}}, \text{Forest_cover_suit}, p_1, \text{Forest_cover_500suit}, p_2, \text{Distance_suit}, p_3, \text{Elevation_suit}, p_4)$$

Where p_{inh} is the probability that the habitat is not suitable even if all the individual factors are suitable, and p_{1-4} are the probabilities that each factor will be required for the habitat to be suitable.

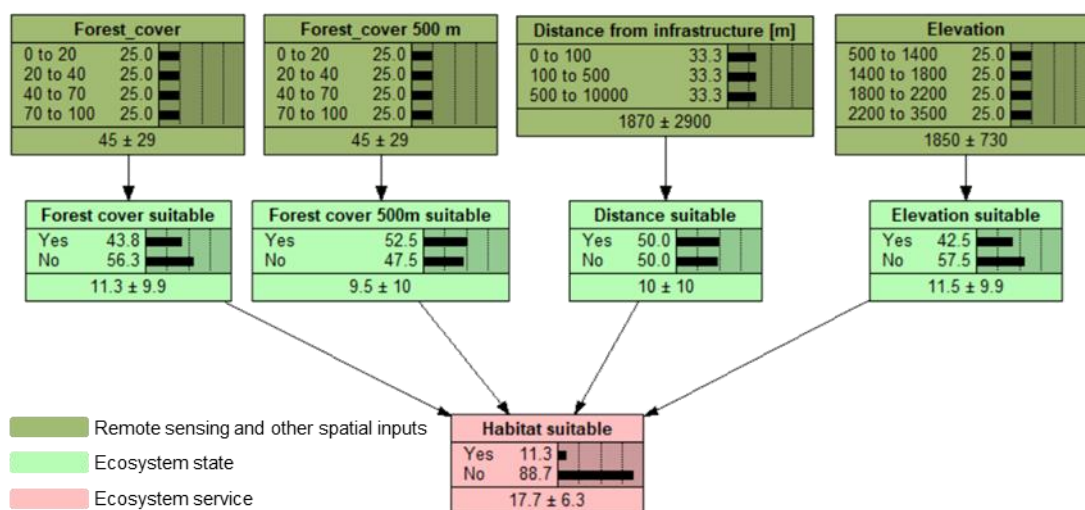


Figure 5.3-1: Bayesian Network describing the habitat suitability for grouse in a mountain region.



6. Case studies

6.1 Avalanche protection in Davos, Switzerland

Authors: Ana Stritih, Peter Bebi, Adrienne Grêt-Regamey

Protection from natural hazards such as avalanches is one of the most important ecosystem services provided by mountain forests. Forests decrease the probability of an avalanche release (Bebi et al. 2009), and reduce the mass and velocity of avalanches that flow through them (Feistl et al. 2014). The capacity of forests to provide avalanche protection depends on other side on their structure and species composition, which can be derived from EO data. On the other side, the demand for avalanche protection depends on the risk to human life and infrastructure, which can also be mapped using remote sensing. While data and models exist for some components of the avalanche protection system, they have not been integrated into a comprehensive model of the ecosystem service, and are associated with large uncertainties. We address this issue by developing a BN, which integrates existing models, EO data, and expert knowledge. The BN is used to more precisely map the ecosystem service and quantify the associated uncertainties (Stritih et al. 2019).

6.1.1 Developing the Bayesian Network

We based our avalanche protection model on previous models developed for this ES (Grêt-Regamey and Straub 2006, Grêt-Regamey et al. 2013) but extended it to incorporate newly available remote sensing inputs as well as recent developments in modelling forest-avalanche interactions. The BN structure was developed through an iterative process of literature review, consultation with experts, and testing the behaviour of the network with different input values.

The input nodes of the network are remote sensing variables, which are proxies for ecosystem structure, and in-situ or modelled avalanche data. These are linked to nodes that describe the natural hazard process, ecosystem functions, and demand for avalanche protection (based on a risk assessment approach).

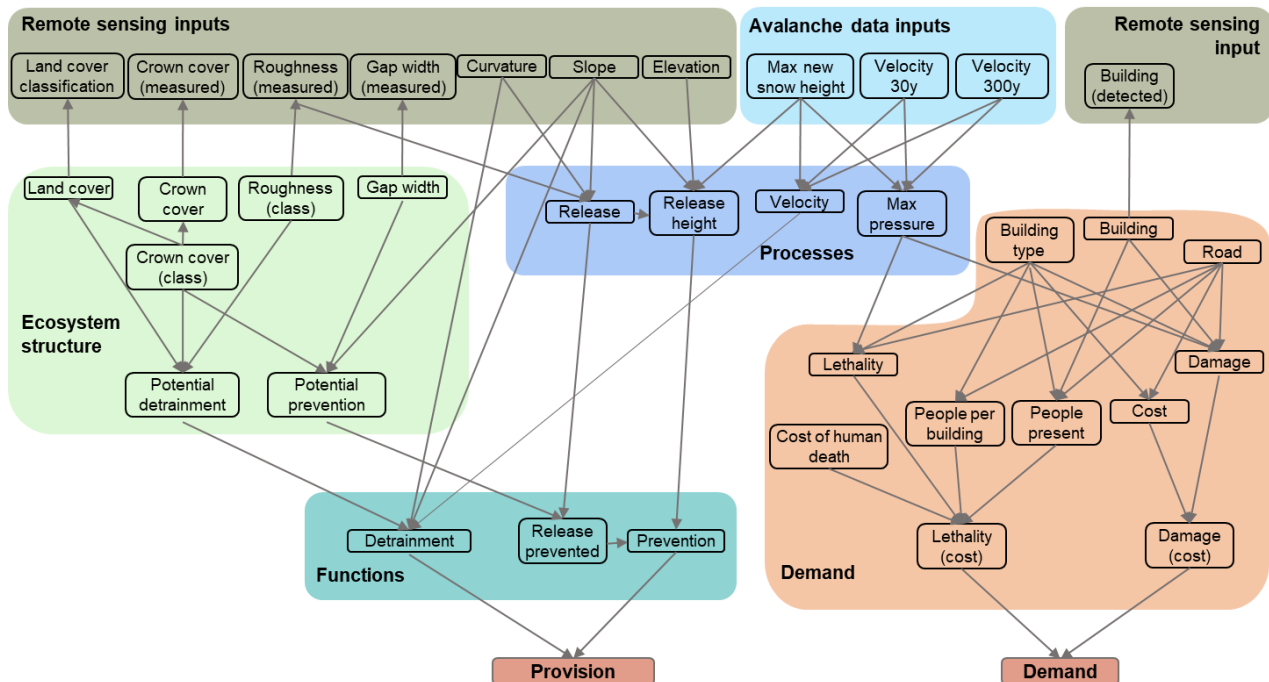


Figure 6.1-1: Bayesian Network developed to model the ES of avalanche protection. The nodes are grouped and coloured based on the types of variables they represent. Spatial inputs are remote sensing and avalanche data, which are linked to variables describing ecosystem structure, avalanche hazard processes, ecosystem functions, and risk factors. The outputs of the network are the provision and demand for avalanche protection. Arrows represent causalities, not the flow of information, and are therefore oriented from ecosystem structure variables to the corresponding remote sensing inputs.

Adapted from (Stritih et al., 2019).

6.1.2 Quantifying the network (CPTs)

In order to integrate different types of information (including remote sensing, process-based models, empirical models, and expert knowledge) and account for the associated uncertainties, we used a variety of methods to quantify the links between nodes in the network.

6.1.2.1 Links between remote sensing proxies and the actual state of the ecosystem

Remote sensing products (e.g. land cover classifications or LiDAR-based measurements of vegetation cover) represent a proxy of the actual state of the ecosystem. We make the uncertainty in the measurements or classifications explicit by creating separate nodes representing the observed value (Y) and the actual state (X) of the variable. The observation is caused by the actual state, not vice-versa, and defining the structure of the network based on this causality helps to define conditional probabilities.

We used this principle to account for uncertainties in the land cover classification. Classification errors are commonly expressed in confusion matrices, which contain counts of predicted classes for objects where the true class is known (in our case, obtained from 110 ground truth locations), with rows representing the classes in reality (c), and columns representing the classes predicted by the classification (c'). Based on these counts, we can calculate either backward probabilities $P(X = c | Y = c')$ (e.g. the probability that a patch classified as forest is a forest in reality); or the forward probabilities $P(Y = c' | X = c)$ (that a forest patch will be classified as forest). The backward probabilities depend on the prior distribution of land cover – if we sample ground truth locations in a densely forested landscape, it is likely that many of the patches classified as forest will in fact be forested, leading to a higher

backward probability than if we sample in a sparsely vegetated area. On the other hand, forward probabilities are inherent to the error process in the remote sensing data and the classification algorithm (Cripps et al. 2009), and are therefore consistent over the whole area. Therefore, we define the classification node Y as the child of the actual class X, and the rows of its CPT then correspond to the forward probabilities $P(Y | X)$.

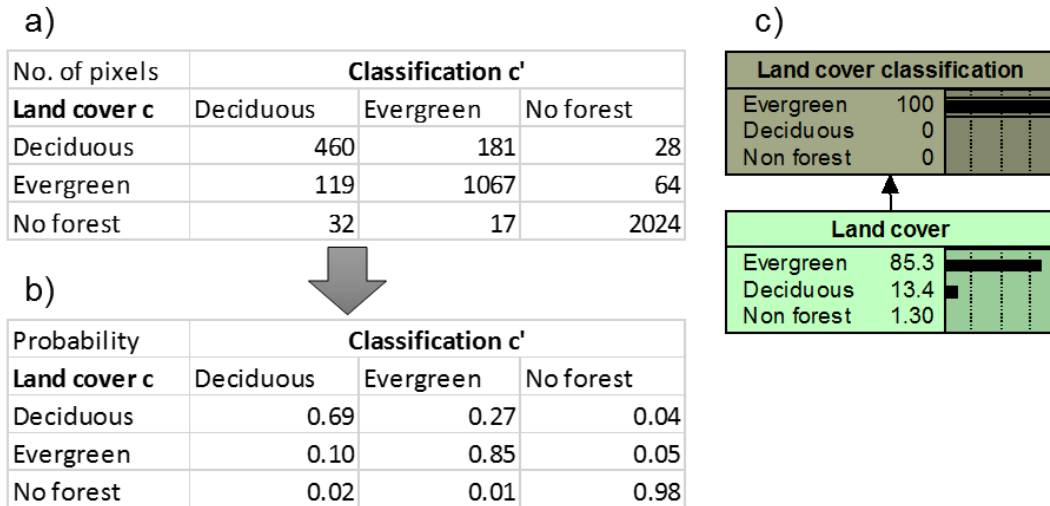


Figure 6.1-2: Confusion matrix of the land cover classification (a), the resulting CPT (b), and the resulting probability distribution when a pixel is classified as evergreen forest (c).

For continuous remote sensing variables (such as the percentage of crown cover) with a known measurement error rate, we similarly define the measured cover Y as a child of the actual cover X. Assuming a normal distribution of errors, we define the conditional probability of Y as a normal distribution $p(Y | X = x) = N(x, \sigma^2)$ where the mean is the value of the actual state (x), and the standard deviation σ is defined by the measurement error. If we have no prior information about the actual state of X, a finding on the child Y (measurement) node then results in a normal distribution $p(X | Y = y) = N(y, \sigma^2)$ of the parent X (actual state).

Since the error of lidar-based crown cover measurements was estimated to be 12 %, the CPT for lidar-measured crown was defined in Netica using the equation:

$$p(\text{Crown_cover_Lidar} | \text{Crown_cover}) = \text{NormalDist}(\text{Crown_cover_Lidar}, \text{Crown_cover}, 0.12 * \text{Crown_cover})$$

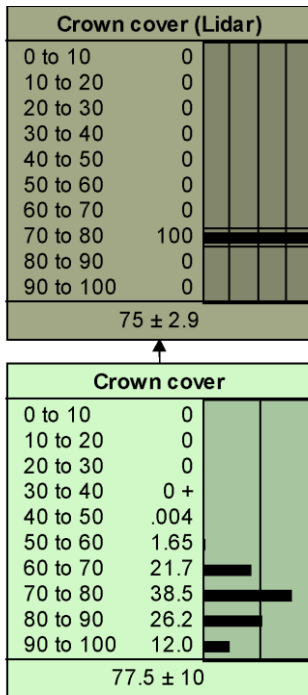


Figure 6.1-3: Distribution of actual crown cover, given a measurement of crown cover. The CPT of Crown cover (Lidar) is defined as a normal distribution around the actual crown cover.

6.1.2.2 Learning from process-based models

The process-based avalanche model RAMMS (Christen et al. 2010) simulates avalanche flows and also snow detrainment in forests during avalanches. In order to quantify the CPT of the node “Detrainment”, we simulated five known avalanche events in RAMMS with varying input parameters (e.g. different snow heights and parameters of snow erodibility, to account for uncertainty in the model). Then, we used the outputs of the simulations to “learn” the CPT of “Detrainment”, using the Expectation Maximisation algorithm in Netica.

6.1.2.3 Incorporating empirical models

Information on forests’ potential to prevent avalanches was available in the form of an empirical logistic model (Bebi et al. 2001). We included this model into the network using a Netica equation, including the model parameter uncertainty by incorporating the parameters not as single values, but as nodes with their probability distribution (normal distributions defined by the parameter estimates B and their standard errors). Then, we calculate the CPT of “Potential prevention” with an equation:

$$p(\text{Potential_prevention} \mid \text{Slope, Gap_width, Crown_cover, Prevention_intercept, cover_B, slope_B, gap_B}) = 1 - (\text{logistic}((\text{Prevention_intercept}) + \text{cover_B} * \text{Crown_cover} + \text{gap_B} * \text{Gap_width} + \text{slope_B} * \text{Slope}))$$

However, this procedure results in a very large CPT for the node (a line for each combination of parameters and predictor variables). Since the parameter nodes will not be modified with evidence, we can reduce the CPT by using the function “Absorb nodes”, which removes the nodes from the network, but retains the associated information in the reduced CPT.

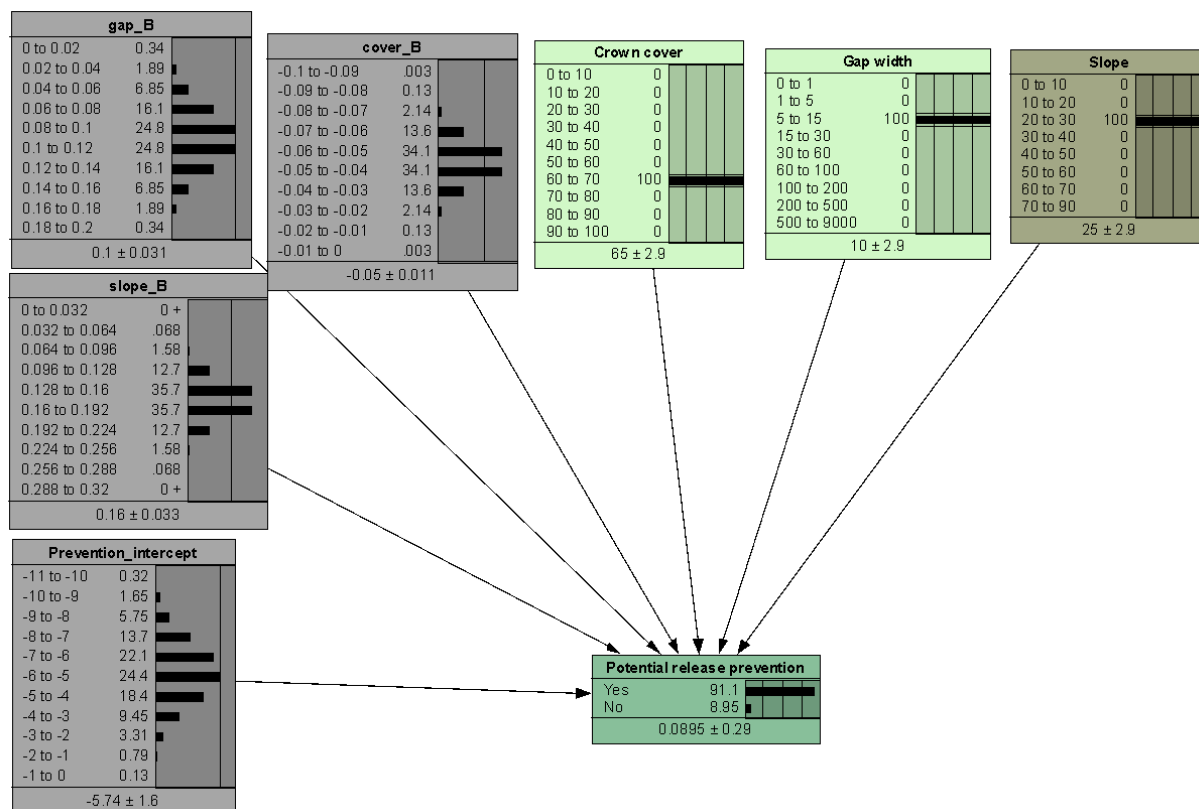


Figure 6.1-4: The parameters of an empirical model can be included explicitly as nodes in the network, to account for model uncertainty when calculating the CPT. Then, these nodes can be "absorbed" to reduce the size of the CPT.

6.1.2.4 Expert knowledge: linking quantitative variables to qualitative categories

Expert knowledge is often related to qualitative categories rather than quantitative variables. For example, it may be easier for an expert to estimate the avalanche protection capacity of forests that are either “open”, “scattered”, or “dense”, rather than based on a percentage of crown cover. Linking such categories to numerical values is associated with a type of linguistic uncertainty (vagueness), where the delineation between categories is not sharp (Regan et al. 2002). Linguistic uncertainty is commonly addressed using fuzzy logic (Zadeh 1965), where membership functions $m(y)$ define the level of membership (between 0 and 1) in a specific class for values of y . For example, we define trapezoidal membership functions of crown cover (Y) for the classes of forest density (X) (see Figure 6.1-5, a). The thresholds between classes have been defined by experts, whereas the slopes of the membership functions are defined based on the standard deviation of measured crown cover at locations where the forest density was classified in the field (method adapted from (Petrou et al. 2013)). At the expert-defined threshold of $Y = 70\%$ crown cover, the probability of the forest being classified as “dense” is 0.5, while a forest with 100% crown cover will certainly be classified as “dense” ($P(X = \text{dense}) = 1$). We use the membership function to define the probability of the class (X) given an observation y , $P(X|Y=y)$, which is proportional to $P(Y|X) * P(X)$.

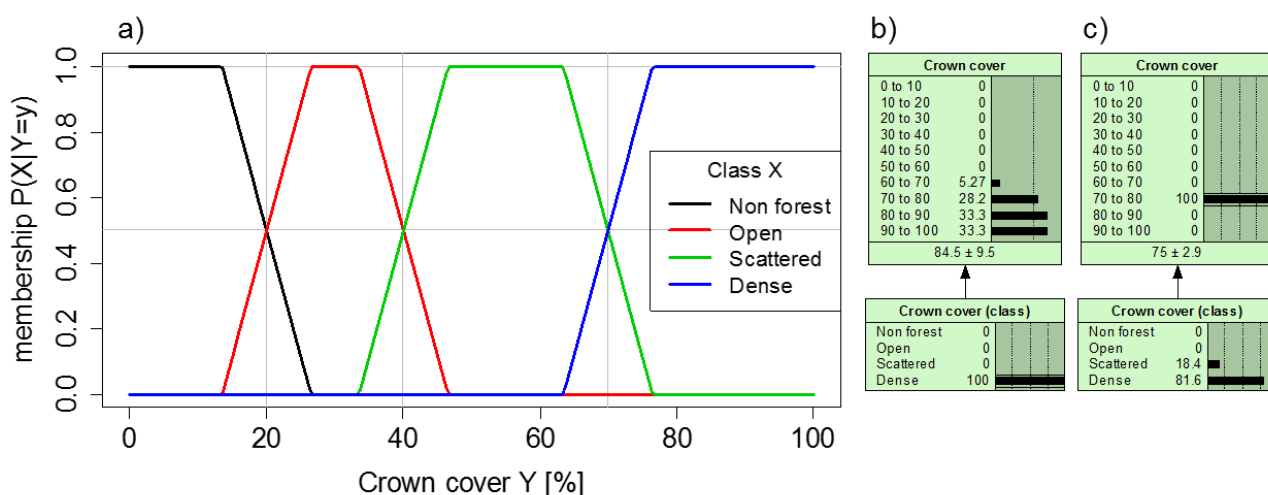


Figure 6.1-5: An illustration of the use of fuzzy logic to translate quantitative variables to qualitative categories:: a) membership functions that define the classes of density based on the percentage of crown cover; b) probability distribution of crown cover for the class “dense”, based on the membership function; c) probability of classes when we observe a crown cover of 70 %.

6.1.2.5 Expert knowledge: estimating distributions

For nodes where no data was available (e.g. “Potential detrainment”), we used expert knowledge to quantify the CPT. To avoid overconfidence, we used the “four-point estimation method” (Speirs-Bridge et al. 2010), where we asked the expert to estimate the lowest and highest value they would expect, the most likely value, and their confidence that the true value is within this range (Metcalf and Wallace 2013). For example, for a dense evergreen forest on rough terrain, the expert estimated the lowest possible detrainment factor to be 24 Pa, the highest 96 Pa, and the best estimate at 48 Pa, with a confidence of 80%. This gives us the quantiles and mode of the distribution, to which we fitted a simple asymmetric triangular distribution (see Figure 6.1-6).

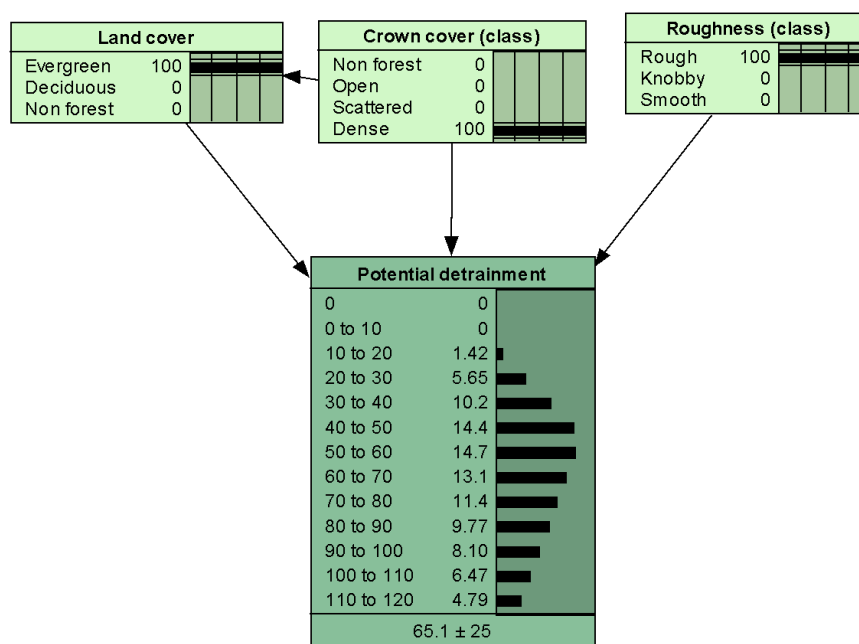


Figure 6.1-6: Expert-based distribution of potential detrainment for a dense evergreen forest on rough terrain.

6.1.3 Spatial application

The spatial inputs to the avalanche protection BN are remote sensing variables, including a land cover classification (derived from a combination of Sentinel2 and aerial LiDAR data) and variables derived from high-resolution aerial LiDAR, such as crown cover, terrain roughness, and detected buildings. In addition, modelled avalanche velocities under two scenarios, extreme (300-year) and frequent (30-year), provide information on the spatial patterns of the avalanche hazard.

Using raster inputs, we performed inference for each pixel in a 5 m resolution raster of the study area. Since the provision and demand for avalanche protection do not occur at the same location, and spatial processes could not be modelled in the pixel-based BN, we quantified provision and demand separately. Thus, we obtained posterior probability distributions of avalanche protection provision and demand for each pixel. In order to map the outputs, we calculated the per-pixel median and entropy (uncertainty) of the posterior probability distributions.

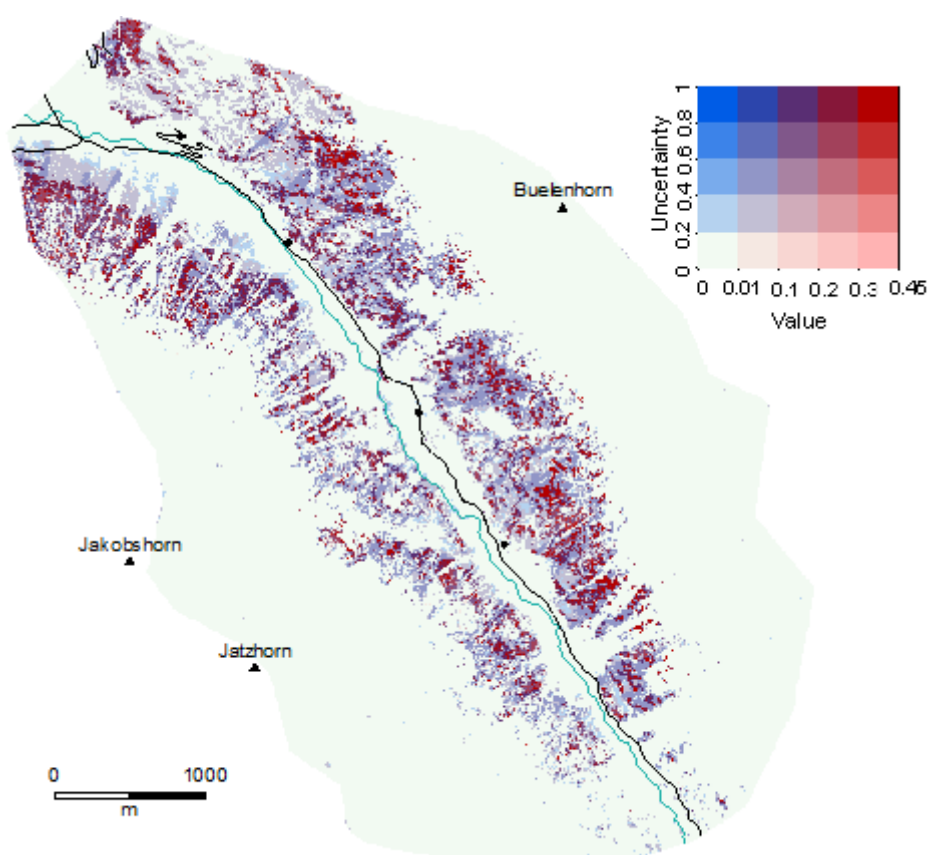


Figure 6.1-7: Modelled provision of avalanche protection in the Dischma valley (5 m resolution). The value is expressed in m of snow, while the uncertainty is calculated as the entropy of the posterior probability distribution. Most areas with a high value of the service also have a high uncertainty (dark red), as do some forested areas with a predicted low protection value (dark blue). Only areas with a zero or very low (light blue) value of the service show a high certainty. From (Stritih et al. 2019).

6.1.4 Validation and sensitivity analysis

The model and the resulting maps of avalanche protection provision and demand, as well as the underlying ecosystem functions, were presented and discussed with local experts. In addition, we performed a sensitivity analysis of the model, using the Netica function “Sensitivity to findings” to calculate the reduction of entropy (uncertainty) on the target nodes in response to findings on other nodes in the network. The entropy reduction

(also called mutual information) gives us an indication of which variables in the system have the highest influence on the ecosystem service.

We also performed a stepwise sensitivity analysis to visualize the flow of information in the network. For each node X, we calculated the proportion of its entropy that can be reduced by a finding on each of its parents. These relative mutual information values were used as weights for links between nodes in a Sankey diagram of the network (Figure 6.1-8) For each node, the thickness of incoming (from the left) links show how much the entropy on the node can be reduced by findings on preceding nodes. Mutual information is not additive, i.e. if both parent nodes can reduce the entropy of a child by 50%, this does not mean that findings on both parents will result in complete certainty on the child node. Nonetheless, plotting the MI gives an indication of the main sources of uncertainty in the model. When the value of MI for all the parents of a node is rather low, this means that the node will have a wide probability distribution even when the states of its parents are known, implying high uncertainty in the corresponding links. If such a node has a large influence on the outcome of the network, this indicates a knowledge gap.

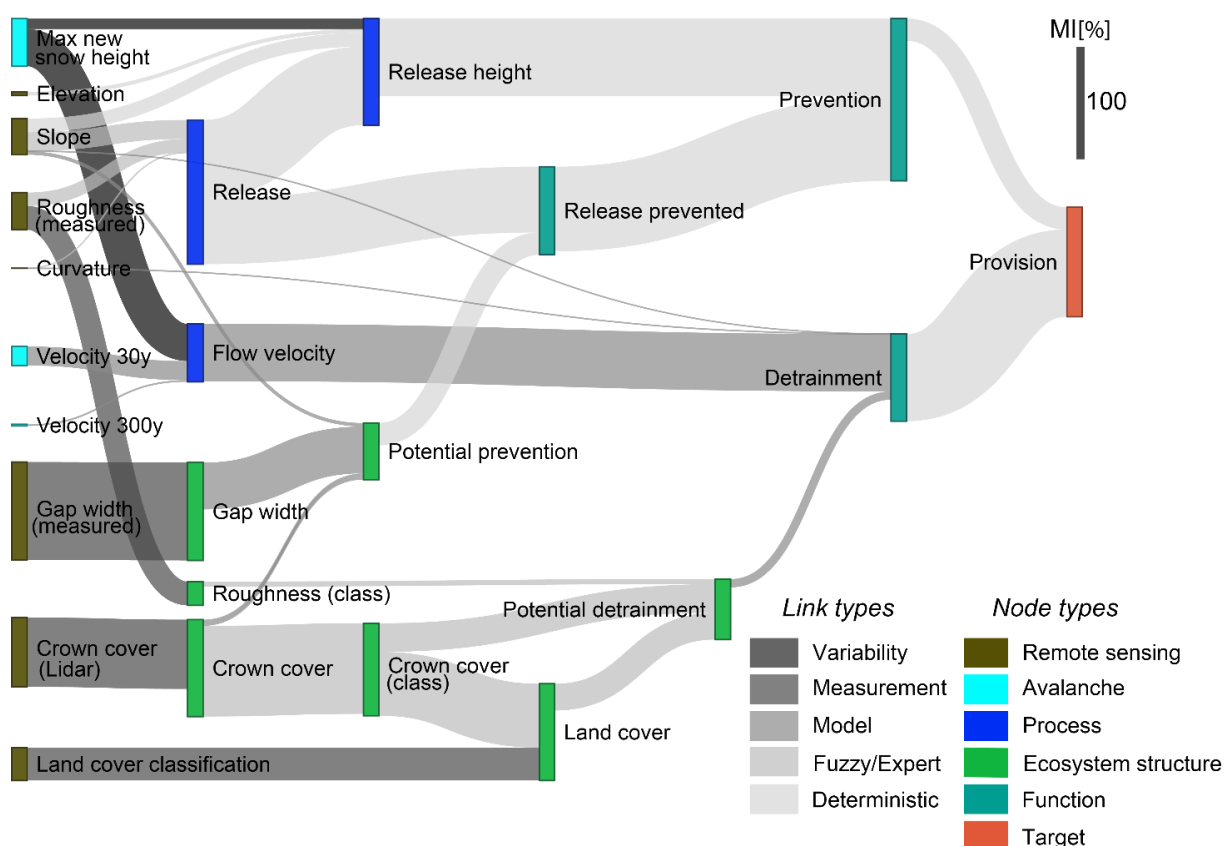


Figure 6.1-8: Stepwise sensitivity analysis of the BN, where the width of a link between two nodes corresponds to the relative mutual information (MI %), i.e. the percentage of the entropy on a node that can be reduced by a finding on a preceding node. The nodes are labelled and coloured by the type of variable represented (see Fig 6.1-1), while the link colours represent the types of uncertainty taken into account while quantifying the link in the BN. From (Stritih et al. 2019).

Overall, the uncertainties related to avalanche processes contribute more to the final uncertainty in ES provision than uncertainties about ecosystem structure. For example, the node “Release” (describing whether a pixel is in a potential avalanche release area) has an important influence on subsequent nodes in the network, but findings on its parents (“Slope”, “Roughness (measured)” and “Curvature”) can only reduce a small part of its entropy, so it is a major source of uncertainty in the model. Some remote sensing inputs have a strong effect on the knowledge



about ecosystem structure (“Gap width” and “Crown cover”), while others have higher uncertainty (e.g. “Roughness”). There is high uncertainty in land cover classification, as its mutual information with actual land cover is only 29 %. However, additional information on actual land cover is gained from the crown cover class (MI = 59 %). The links from ecosystem structure to the potential provision of ES also contain high uncertainty, regarding both the potential of a forest to prevent avalanches (empirical model-based “Potential prevention”) and to stop snow during an avalanche (expert-based “Potential detrainment”). However, “Potential detrainment” has a relatively low influence on the corresponding ecosystem function (process model-based “Detrainment”). This function is affected more strongly by the avalanche process (“Velocity”), which in turn is affected by the natural variability in release conditions (“Max new snow height”).

6.1.5 Description of the network nodes

Table 6.1-1: Description of nodes in the avalanche protection Bayesian Network, their states, and methods use to quantify their conditional probabilities.

Node	Description	States	CPT method
Input nodes - remote sensing			
Land cover classification	Random forest classification with combination of LiDAR, aerial CIR images and Sentinel2	Evergreen forest, deciduous forest, non-forest	Confusion matrix
Crown cover (measured)	Measured cover of vegetation above 3 m	10 states: 0 - 100 [%]	Normal distribution around actual state
Roughness (measured)	Measured terrain roughness, derived from LiDAR-based DTM (Sappington et al., 2007)	9 states: 0 - 1	Fuzzy definition of categories based on ground truth
Gap width (measured)	Width (along contour lines) of non-forested area [m], derived from LiDAR-based CHM	9 states: 0 - 8000 [m]	Normal distribution around actual state
Building (detected)	Presence of a building detected from LiDAR	Boolean	Confusion matrix
Curvature	Planar slope curvature, derived from LiDAR-based DTM	7 states: -45 - 5	
Slope	Slope angle, derived from LiDAR-based DTM	8 states: 0 - 90 [°]	
Elevation	Elevation [m a.s.l.]	7 states: 1500 - 2900	
Input nodes - avalanche hazard data			
Max new snow height	Annual maximum new snow height, which determines the avalanche release scenario	9 states: 0 - 2 [m]	Gumbel distribution of maximum new snow height, Davos, Fluelastrasse



Velocity 300	Maximum velocity under the 300 year scenario, modelled in RAMMS	9 states: 0 - 60 [m/s]
Velocity 30	Maximum velocity under the 30 year scenario, modelled in RAMMS	10 states: 0 - 60 [m/s]

Nodes representing ecosystem structure

Crown cover	Cover of vegetation above 3 m	10 states: 0 - 100 [%]	Fuzzy logic definition of categories
Crown cover (class)	Category of forest density	Dense, scattered, open, non-forest	
Roughness (class)	Category of terrain roughness	Rough, knobby, smooth	
Land cover	Actual land cover class	Evergreen forest, deciduous forest, non-forest	Based on ground truth plots
Gap width	Width (along contour lines) of non-forested area [m]	9 states: 0 - 8000 [m]	
Potential release prevention	Potential of forest to prevent an avalanche release	Boolean	Logistic mode (Bebi et al., 2001)
Potential detrainment	Capacity of forest to remove snow from avalanche flow	12 states: 0 - 120 [Pa]	Expert knowledge (four-point estimation)

Nodes representing the hazard process

Release	Probability of an avalanche release	Boolean	Fuzzy logic
Release height	Height of snow released in the event of an avalanche	10 states: 0 - 3 [m]	Logical combination of "Release" and "Max new snow height", corrected for "Slope"
Velocity	Maximum avalanche flow velocity	9 states: 0 - 60 [m/s]	Defined by scenarios, with uncertainty estimated based on simulations with varying input parameters
Max pressure	Maximum avalanche pressure	4 states: = 0 - 300 [kPa]	Derived from velocity, with uncertainty estimated based on simulations with varying input parameters

Nodes representing ecosystem functions

Release prevented	Probability a release would be prevented by the forest	Boolean	Logical combination of "Release" and "Potential prevention"
-------------------	--	---------	---



Prevention	Height of snow in prevented avalanche release	11 states: 0 - 3 [m]	Logical combination of "Release prevented" and "Release height"
Detrainment	Height of snow detrainment in forest during avalanche	7 states: 0 - 0.8 [m]	Learning from RAMMS simulations with varying input parameters

Risk assessment and valuation nodes

Building	Presence of a building	Boolean	
Building type	Type of building	One-family house, farm building, guesthouse, multi-family house, industrial, infrastructure	Distribution of types from local data
Lethality	Avalanche pressure is lethal	Boolean	Fuzzy logic based on values from literature (Swiss National Platform for Natural Hazards)
People per building	Average no. of people per building	6 states: 0 - 20	Fuzzy logic based on values from literature
People present	People are present in a building	Boolean	Fuzzy logic based on values from literature
Damage	Building is destroyed by avalanche	Boolean	Fuzzy logic based on values from literature
Building cost	Cost of destroyed building	6 states: 0 - 10 ⁷ [CHF]	Distribution per type from local data
Damage (cost)	Cost due to damaged building	7 states: 0 - 10 ⁷ [CHF]	Logical combination of "Damage" and "Building cost"
Cost of human death		Constant: 5*10 ⁶ [CHF]	Constant value from literature, Life Quality Index approach (Merz et al. 1995)
Lethality (cost)		5 states: 0 - 10 ⁸ [CHF]	Logical combination of "Lethality" and "Cost of human death"

Output nodes

Provision	Combination of prevented release and detrainment snow height	12 states: 0 - 4 [m]	Sum of "Prevention" and "Detrainment"
Demand	Avalanche risk to people and buildings	5 states: 0 - 1.1*10 ⁸ [CHF]	Sum of "Damage (cost)" and "Lethality (cost)"

6.2 Land-use scenarios in Sierra Nevada

Authors: Moreno-Llorca R., Ros-Candeira A., Alcaraz-Segura D., Herrero-Lantarón J., Bonet-García F.J., Millares-Valenzuela A. (UGR)

Due to the combined effects of climate change and shifts in land use, the distribution and structure of the vegetation of the Sierra Nevada has been undergoing rapid change, which in turn affects the associated ecosystem services. The cover of tree formations in Sierra Nevada has expanded from 15% to 51.23% over the last 60 years, while the areas of scattered tree cover and natural forests have densified, and the area of cultivated fields has declined (from 17.8% to 4.72%) (Zamora et al. 2016). Therefore, it is important to ascertain future land-use change, as well as its effects on vegetation and ecosystem services.

The main purpose of this study is to facilitate the land-use management of Protected Areas (PAs) based on ecosystem services (ES). A BBN is being designed to develop future land-use scenarios for the Sierra Nevada under different environmental and management conditions. Afterwards, we will implement these scenarios in other ES assessment models. The analysis of ES trade-offs in several scenarios will help managers to predict the state of ES and their relations in the future.



Figure 6.2-1: Framework of the study on future ecosystem services in the Sierra Nevada.

6.2.1 Development of the Bayesian Network

The process to develop the BBN began with a literature review of several models of land use change (Renwick and Revoredo-Giha 2008, Díaz et al. 2011, Lamarque et al. 2013, Celio et al. 2014, Celio and Grêt-Regamey 2016) with a special emphasis on the actors who make the decision to change land use. Once a preliminary set of drivers of land use change was defined, we used an iterative process of consultation with experts and workshops with stakeholders to define the nodes of the network, their different states and the causal relationships between variables in the BBN.

Interviews were initially carried out with expert researchers on changes in land use and agricultural activity in the Sierra Nevada. In the interviews, the drivers found in the literature were discussed. It was analyzed which drivers

were relevant for the Sierra Nevada case. An initial structure of the Bayesian network was discussed, without addressing the states of the variables.

Afterwards, three workshops were carried out with experts and stakeholders, one in each differentiated zone of Sierra Nevada: north-west, south, and east. Stakeholders in attendance were cattle ranchers, farmers and irrigation organisations. Participating experts were rural development agents and technical staff from the local agricultural offices.

The first activity was to validate the drivers of the part of the network related to stakeholders’ decisions. To do this, the drivers were presented one by one. It was discussed whether this driver was relevant in Sierra Nevada and why. There were drivers that aroused a wide consensus from the beginning. In other cases, the reasons pro or con the importance of the driver were discussed. The final decision to include or exclude a driver could always be made by consensus. Subsequently, the same procedure was used to establish relations between the variables, designating, on a case-by-case basis, the child nodes and the parent nodes.

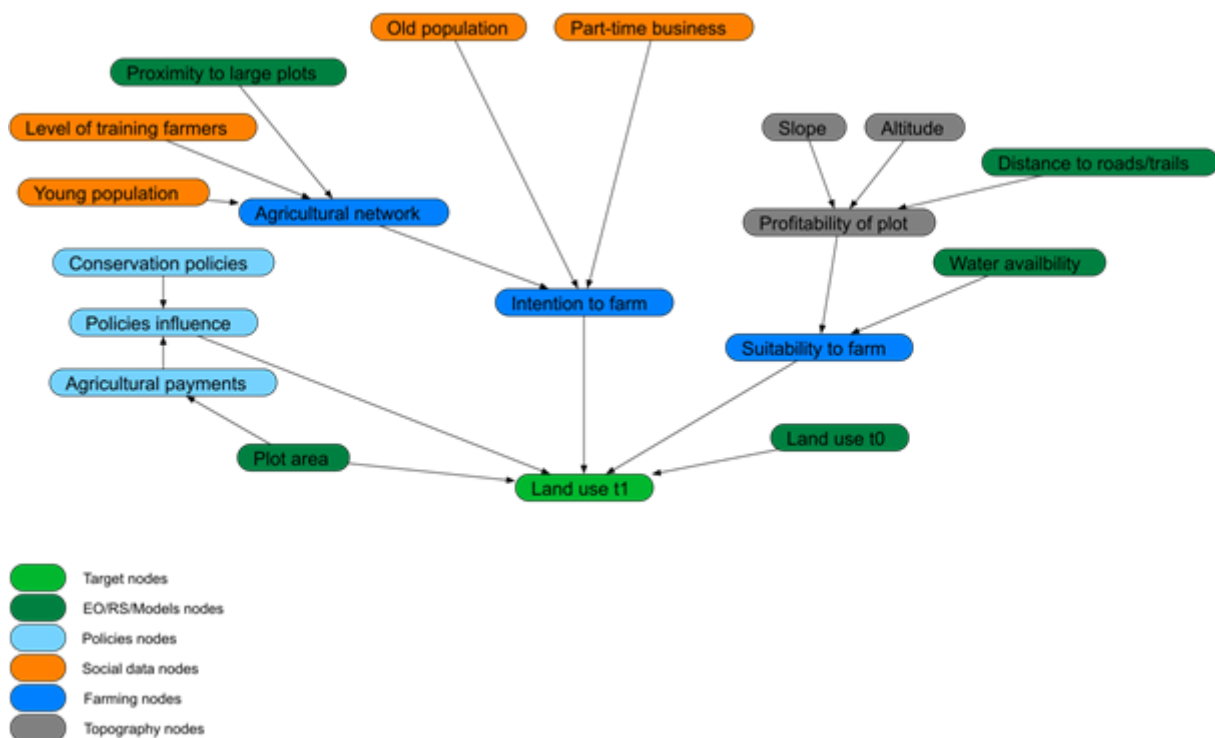


Figure 6.2-2: Bayesian Network developed to model the land use changes in Sierra Nevada (Spain). The nodes are grouped and coloured based on the types of variables they represent. Spatial inputs are EO/RS/Model and topography nodes, while other inputs include policy and social data.

6.2.2 Quantifying the network (CPTs)

After designing the BBN, we planned to parameterize the CPT of the network based on two different type of sources: experts and stakeholders. Due to the preliminary attempts to estimate the probabilities with each type of group, we decided to use separate participation techniques for the experts and stakeholders, to better adapt to each profile's way of understanding probabilities. The part of the network related to stakeholder’s decisions is

parameterized with local people and Sierra Nevada PA-staff. All other variables will be parameterized by expert knowledge.

6.2.2.1 Stakeholder knowledge

While experts are more familiar with the concept of probability, it is difficult for stakeholders to decide in terms of the probability of a state as a function of a combination of states of the parent nodes in a network. They are more accustomed to choosing which is the most suitable (more likely) choice between two different options. Therefore, we used a choice-experiment approach, where the stakeholders are presented with two options based on combinations of parent node states, and they choose which option is suitable.

To this end, we designed an online survey for each child node of the network. The survey iteratively asks the respondent to choose between two scenarios based on different combinations of the states of the parent variables. As the respondent is replying, the software detects the missing combinations of the variables' states to have sufficient answers. Once the software has obtained information about all combinations (this is configurable to a specific set of information), the survey ends.

• To define the variables
• To establish the states of the variables
• To determine the causal relations

	Option A	Option B	No choice
Type of crop	Dry woddy crop	Irrigated horticultural crop	
Slope	Lower 12 percent	Over 25 percent	
Altitude	Higher 1400 mas	Lower 1400 mas	
Access to the plot	Adjoin to road	Out of road	
check the best option	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

CPTs

Figure 6.2-3: Workshops with stakeholders were carried out to define variables, the states of variables and to determine causal relations. Afterwards, individual surveys, based on scenarios selection, were carried out with stakeholders to parameterize the CPTs.

The results of the survey are converted to a frequency table, which is used to fill the CPT of the corresponding node.



Table 6.2-1: Example of the parameterisation of a CPT based on the survey, for one of the combination of profitability of crop node.

	Slope	Over 25%			
	Altitude	Lower 1400 mas		Over 1400 mas	
	Access	Next to a road	Far from road	Next to a road	Far from road
Profitability of plot	High	0,56	0,15	0,36	0,01
	Low	0,44	0,85	0,64	0,99

6.2.2.2 Expert knowledge

We will carry out two workshops with experts, who have a better understanding of the approach of defining probabilities for the different states of the children variables. The experts were selected on the basis of scientifically relevant criteria to represent the following categories: ecology, agriculture, livestock, hydrology, landscape and land-use and finally, conservation and agricultural management.

6.2.3 Spatial application

The BBN is developed to assess future land use scenario in a spatial way. There are three types of spatial nodes in the BBN:

- Target node. Land use on time 0 and time 1 are the main spatial nodes, where land use at time 0 is an input to the model, while land use at time 1 is the output.
- EO/RS/Model nodes. Several nodes are spatial information from Earth Observation/Remote Sensing/Model outputs: proximity to large plot, plot area, distance to roads and water availability.
- Topography nodes. Slope and altitude are topographical variables used as input in the BBN.

The spatial information come from different sources and formats. Raster information has 10 meters' resolution while vector information is on a scale of 1:25,000. The extent of the study is the limit of Natural and National park of Sierra Nevada.

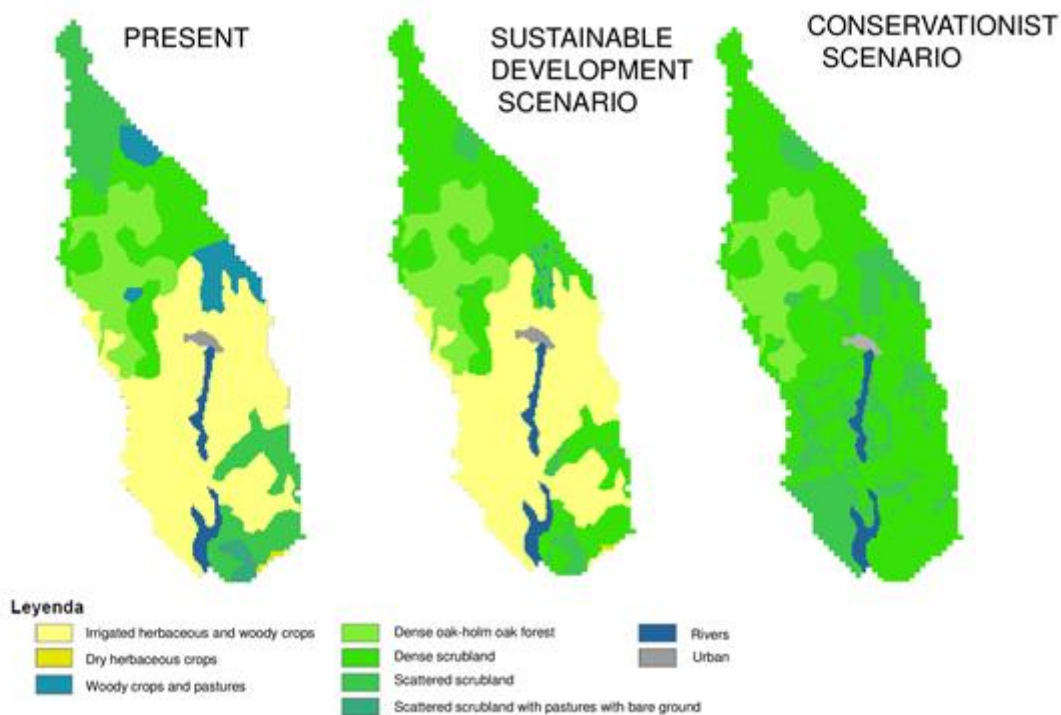


Figure 6.2-4: Modelled land uses scenarios using BN-spatial R package. Preliminary implementation of Sierra Nevada BBN in a watershed of Nevada municipality. The map shows the most likely class of land use.

6.2.4 Dynamics

We are planning to run the model in a dynamic approach in the implementation of the BBN in gBay. Land use at time 0 is the first input in the network and land use at time 1 is the result of the first step. Afterwards, land use at time 1 is the input in step 2 and so on.

The temporal resolution is five years and the time period to analyze is fifty years.

6.2.5 Description of the network nodes

Table 6.2-2: Description of nodes in the land use change network, their states, and methods use to quantify their conditional probabilities.

	Description	States	CPT method
Input nodes - Topography nodes			
Slope	Slope percent-based DTM. GIS analysis.	3 states: <ul style="list-style-type: none"> · Below 12 percent · Between 12 and 25 percent · Above 25 percent 	
Altitude	Elevation of a plot [m.a.s.]. GIS analysis.	2 states: <ul style="list-style-type: none"> · Below 1400 m.a.s. · Above 1400 m.a.s. 	
Input nodes - EO/RS/Model nodes			



Distance to roads	The distance between a plot and a road or a trail to facilitate the harvesting. GIS analysis.	2 states:	<ul style="list-style-type: none"> · Next to road · Far from road
Plot area	Size of the plot. GIS analysis.	3 states:	<ul style="list-style-type: none"> · Smaller than 3 has. · Between 3 and 10 has. · Greater than 10 has.
Water availability	Water availability to irrigation	3 states:	<ul style="list-style-type: none"> · Low · Medium · High
Proximity to large production plots	Proximity to large production areas	2 states:	<ul style="list-style-type: none"> · In the same municipality · In other municipality
Land use t0	Reclassification of government land use-land cover map.	8 states:	<ul style="list-style-type: none"> · Urban · Dry woody crops · Irrigated horticultural crops · Irrigated woody crops · Pastures · Scrubland · Gallery forest · Forest

Input nodes - Social data nodes

Part time business	Opportunity to complement agricultural production with other economic activity	2 states:	<ul style="list-style-type: none"> · Yes · No
Old population	Presence of population over 65 years of age	3 states:	<ul style="list-style-type: none"> · Lower 10 Percent · Between 10 and 20 Percent · Higher than 20 percent
Level of training	Level of training and knowledge of farmers	2 states:	<ul style="list-style-type: none"> · Low · High
Young population	Presence of population under 40 years of age	3 states:	<ul style="list-style-type: none"> · Lower 10 Percent · Between 10 and 20 Percent · Higher than 20 percent

Input nodes - Policies nodes



Conservation policies	Permissiveness agricultural activities	with 3 states:	<ul style="list-style-type: none"> · Low · Medium · High
Agricultural payments	Right to receive agricultural payments	2 states:	<ul style="list-style-type: none"> · Right to payments · No right to payments

Nodes representing farming behaviour

Agricultural network	Social network of farmers	? to be defined	Expert workshops
Intention farm	Intention to engage in agricultural activity	? to be defined	Expert workshops

Node representing policies influence

Policies influence	Influence of the environmental and agrarian policies in the decision of cultivating or not a certain place.	3 states:	<ul style="list-style-type: none"> · Pro abandonment · Neutral · Supporting traditional use 	Expert workshops
--------------------	---	-----------	--	------------------

Node representing profitability of plot

Profitability of plot	Suitability of a particular site for cultivation based on biophysical criteria	3 states:	<ul style="list-style-type: none"> · Low · Medium · High 	Stakeholder surveys
Suitability farm	Suitability of a particular site for cultivation based on biophysical criteria and water availability	2 states:	<ul style="list-style-type: none"> · Low · High 	Stakeholder surveys



6.3 Collaborative modelling of ecosystem services in Doñana

Authors: Pablo F. Méndez, Luis Santamaría (EBD-CSIC)

The overall goal of the modelling exercise is to better understand and predict the influence of multiple environmental and management/policy factors over the waterbird community of the protected marsh of the Doñana PA. The main interests are twofold. First, the interest lies in the ability of BNs to capture the perceptions and beliefs of different stakeholders (e.g., decision-makers vs experts) before and after existing and new information is incorporated into the BN for the modelling of new scenarios. In addition, the interest lies in the use and overall performance of BNs as a tool for ecological prediction and decision support, able to incorporate and update information from different sources, including extra expert knowledge and empirical sub-models, in a modular fashion. In this case, we considered their twofold diagnostic (bottom-up reasoning) and explanatory/predictive (top down reasoning) capability, which allows the modelling process to progress in situations of incomplete information (e.g., missing or still incipient data on key system variables) (Marcot et al. 2006, Castelletti and Soncini-Sessa 2007). We were also interested in their ability to integrate both qualitative and quantitative information (Chen and Pollino 2012). In both cases, we chose BNs due to their appeal for collaborative model development, visualization of causal relations, filling of data gaps and identification of uncertainties (Castelletti and Soncini-Sessa 2007, Kleemann et al. 2017).

Doñana PA's BN model will be developed in two stages:

Stage 1: elicitation of CPTs based on expert judgement.

Stage 2: mapping ecosystem services using information derived from remote sensing and empirical models outputs.

BN development is currently towards the end of Stage 1, which has consisted of different tasks, as described below.

6.3.1 Stakeholder engagement

Here, by stakeholder engagement we refer to the overall process of stakeholder analysis and identification purpose of expert judgement elicitation and their further engagement along the modelling process. This process has benefitted from our long-term research program in Doñana based at the Doñana Biological Station, which has allowed us to develop an extensive collaboration network with managers, researchers and key stakeholders onsite (Méndez et al. 2010, Mendez et al. 2012). Based on purposive, judgement sampling, stakeholders were selected at three different levels: decision-makers (parties with the power of making decisions among alternatives affecting the course of implementations), key stakeholders (parties affected by decisions at an executive level but without direct decision power) and experts (individuals knowledgeable about the factors and issues under investigation, due to their education, training and experience). All of them were considered as model end users. In engaging stakeholders at an early stage we aimed at enhancing the acceptance and trust of model structure and specifications, as well as on final decisions and management actions, through the strengthening of their sense of ownership of both the model's outputs and the decision process informed by them. We then proceeded to design the BN network.



6.3.2 Developing the Bayesian Network

To build the BN network, which we formally call the model’s Directed Acyclic Graph (DAG), we first developed a conceptual model of our study system during a first workshop (1st Stakeholders Workshop, April, 2017) involving key decision makers, general stakeholders and experts. At the outset of the workshop, to trigger initial discussion with participants, we presented recently published statistical analyses published in the context of ECOPotential that identify a set of environmental predictors of waterbird distribution and habitat use in Mediterranean wetlands under climate change, including the Doñana marshes (Ramírez et al. 2017). Then, we held a structured discussion about the criteria that Doñana’s waterbird community must fulfil to be considered in a “good conservation status”. We focused on the richness/diversity of the community and the population trends of common, threatened and representative/indicator species. Finally, we built the conceptual model asking participants to identify factors (i.e. variables), either environmental (e.g., salinity) or social (e.g., management interventions, policies), that would ensure the accomplishment of a final goal established in the prior activity: a stable, diverse and representative waterbird community in Doñana. In a backcasting fashion, participants identified from primary factors directly affecting the final goal (primary causal variables) to secondary factors affecting the primary ones (secondary causal variables), and so on. The process consisted of subsequent participation rounds that ensured the concurrent intervention of all participants (round robins), repeated until everyone considered the model (Figure 6.3-1) saturated.

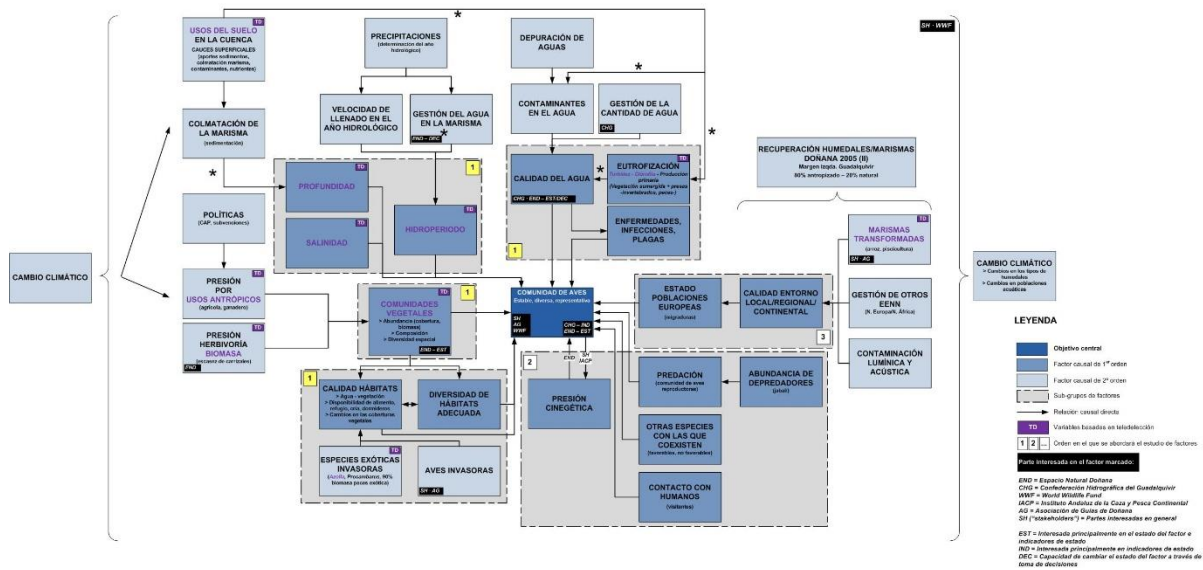


Figure 6.3-1: Conceptual model of our study system developed during the 1st Stakeholders Workshop.

Second, we built a preliminary DAG of the study system in laboratory, based on the workshop’s conceptual model, including only a subset of environmental input variables as parent nodes. The selection of input variables was based on the opinion obtained from key experts at the previous workshop, and the existence of empirical data for the environmental variables so as to allow for future model validation, testing and use for predictive or diagnostic purposes. It was also aimed to assess potential shortcomings of the future BN model derived from data limitations or incorrect variables or interactions, as well as to reduce its complexity (i.e. enhancing its parsimony) thus easing the ascription of probabilities during its parametrization using expert judgement (see below). Nodes were classified using Cain’s (2001, pp. 20-21) (Cain 2001) categories for generalizing network structure, in order to distinguish among decision, random and utility nodes. We used as few states as possible to keep resulting CPTs tractable (Marcot et al. 2006).

In a third stage, we defined our ecological output variables of interest in the DAG. We classified waterbird species in different groups based on the statistical analyses mentioned above. After that, we consulted with two in-house experts as a means of validation of our classification. Based on the final classification obtained, we devised and introduced into the DAG the ecological variables as an output child node (see 6.3.3 below for further details).

6.3.3 Quantifying the network (CPTs)

Finally, we ran a second workshop (2nd Stakeholders Workshop, October, 2018) with a larger group of decision makers, general stakeholders and experts. Our main aims were:

1. To validate: (1) our grouping of breeding and wintering waterbirds; (2) our study system's DAG (Figure 6.3-2), which contains a subset of environmental factors as causal parent variables, obtained in the 1st Stakeholders Workshop.
2. To parametrize the DAG by eliciting expert judgement from the participants in order to: (1) build a first version of our Bayesian Network model; (2) assessing differences in perceptions among the participants (decision makers, key stakeholders, experts), as well as an overall degree of uncertainty from their subjective judgement; (3) assessing the requirements, interoperability and complementarity of this first version with a second-iteration version including empirical information from different sources (e.g., in-situ data, remote sensing data, model outputs) to be used for predictive and/or diagnostic purposes.

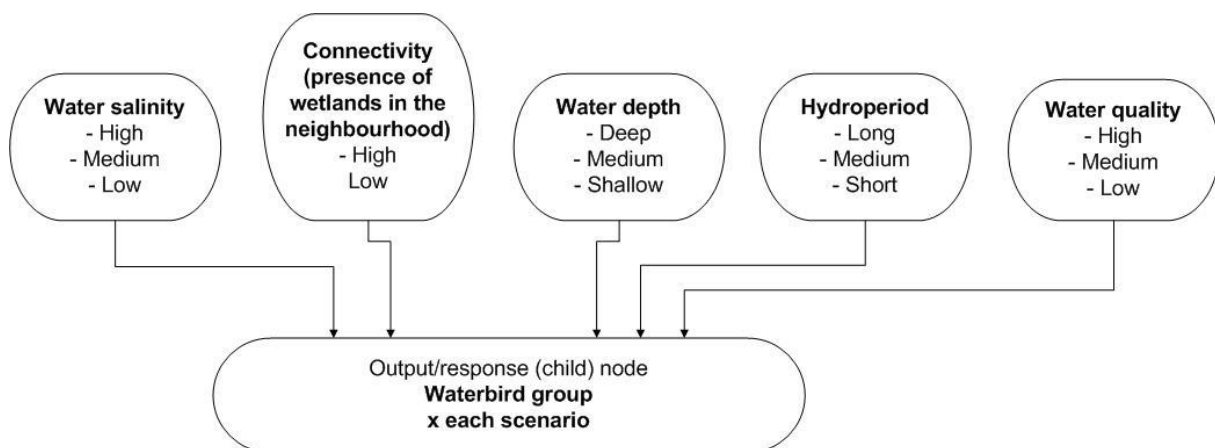


Figure 6.3-2: Directed Acyclic Graph of our study system.

At the outset of the workshop, we described the whole activity, including a short introduction to ECO-POTENTIAL, the results of the first workshop and the objectives and tasks of the second workshop. Then using pre-defined document, we asked participants to rate three different criteria for the groups stemming from our own grouping of the waterbird community, based on the above mentioned statistical analyses: representativeness, conservation value, diversity. This activity was performed for both wintering and breeding species. We provided participants with imagery of the groups from our own classification on screen. We then proceeded to an expert judgement elicitation exercise, asking participants to assess the conditional probabilities linking the DAG's (parent) environmental variable nodes with the occurrence of the events defined by the output response variable (child) nodes. Parent variable nodes were presented as scenarios affecting the output child nodes, consisting of different combination of states of the former nodes. As it was expected that not all participants were at ease with probability reasoning, we provided sheets containing the different scenarios and a verbal/numerical scale to assess their probabilities



over the output nodes. To facilitate the cognitive process and ease the mobilization of the participant's expertise, they were able to use the imagery cards previously provided, which included one representative species for each waterbird group.

Using the verbal/numerical scale provided, the general question to participants was: given this scenario (certain combination of states of the parent nodes), what is the probability that this output is present. Participants were asked to think scenarios as single events and to assume waterbird (separately for wintering and for breeding) populations as already established in the Doñana marshes.

To reduce the number of elicited response probabilities, we used the Conditional Probability Table (CPT) calculator provided by Cain (2001), which uses the probabilities to key scenarios and interpolates all other combinations. This procedure reduced the number of probabilities directly elicited by participants, thus enhancing the consistency of the elicited values and the efficiency of the elicitation process. The CPT calculator reduces the number of scenarios to key anchoring points in a so-called Elicited Probability Table (EPT), which are then interpolated to complete the entire table. In general, key anchoring points consist of: (1) all of the parent nodes set in an extreme state; (2) all of the parent nodes set in the other extreme state; (3) each parent is switched from one extreme to the other consecutively, to set the rest of key anchoring points.

As a final step, a questionnaire was administered to characterize the experience of the participants in the study system. Questions asked were:

1. Name and surname.
2. Main area of work (ecologist, conservationist, business owner).
3. Years of experience or working in connection with the study system in general (e.g., waterbird ecology or conservation, bird watching, hunting, etc.).
4. Years of experience or working in connection with the study system in Doñana.

6.4 The impact of aquatic ecosystems provisioning services on tourism in the Danube Delta

Authors: Constantin Cazacu, Cristian Mihai Adamescu (UB)

The Danube Delta is the second largest delta in Europe, after that of the Volga River and one of the best preserved in the world. The geomorphology of the area is formed by the dynamics of the river arms, low altitudes and the presence of the sea, creating a complex landscape of freshwater ecosystems (canals, shallow lakes, and wetlands), flood plains, alluvial forests, reed-beds, lagoons and coastal area. Due to its high biodiversity and uniqueness of landscapes, the delta attracts about 150000 tourists every year, which is ten times the number of inhabitants.

We developed a storyline aimed at assessing trade-offs between ecosystem provisioning services and tourism and recreation in the delta. System productivity is reflected by water quality and biodiversity, which may influence the tourists' distribution and interests within the area.

6.4.1 Development of the Bayesian Network

Our model is based on the experts' views and consultation with stakeholders' aimed at describing the relationships between the provisioning services and the tourism and recreation activities as cultural ecosystem services provided by the ecosystems from the Danube Delta Biosphere Reserve

The development of the BBN involved a several steps, as we have to accommodate scientific data from observations as well as stakeholders' perceptions.

In a first step, we develop a diagram integrating the essential variables described in the storyline and potential indicators or models that can be used to assess them (Figure 6.4-1).

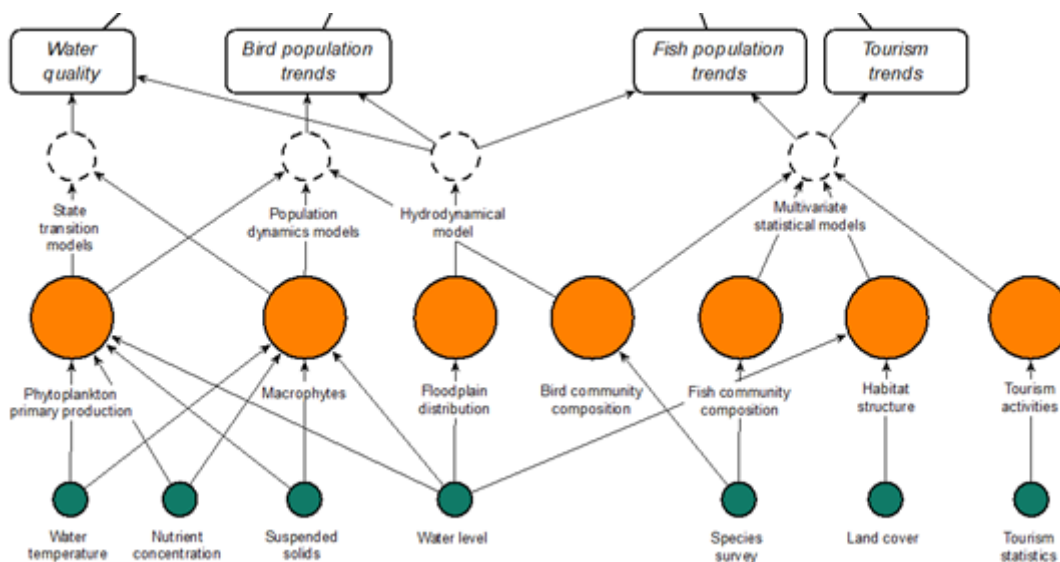


Figure 6.4-1: Diagram representing the storyline developed for the Danube Delta, including the identification of essential variables at the storyline level.

The second step is to develop a BBN, focusing on exploring the link between aquatic ecosystem provisioning services and tourism and recreational activities. We built a network considering several variables that influence the quality/productivity of aquatic ecosystems (e.g. nitrogen phosphorous ratio, total suspended solids, phytoplankton, fish species, macrophytes, and benthic invertebrates) and tourist attractions (e.g. habitat quality, fish productivity, bird density and accessibility) (Figure 6.4-2). Together with stakeholders (mainly researchers), we developed a

graphical representation showing causality relationships between the variables that are influencing the system productivity and the recorded distribution of tourists.

Three categories of stakeholders were involved in the process: researchers specialized in ecology, ornithology, fishery and chemistry (from the University of Bucharest and from Danube Delta National Institute), managers from Danube Delta Biosphere Reserve administration and representatives of local companies dealing with tourism and commercial fishery.

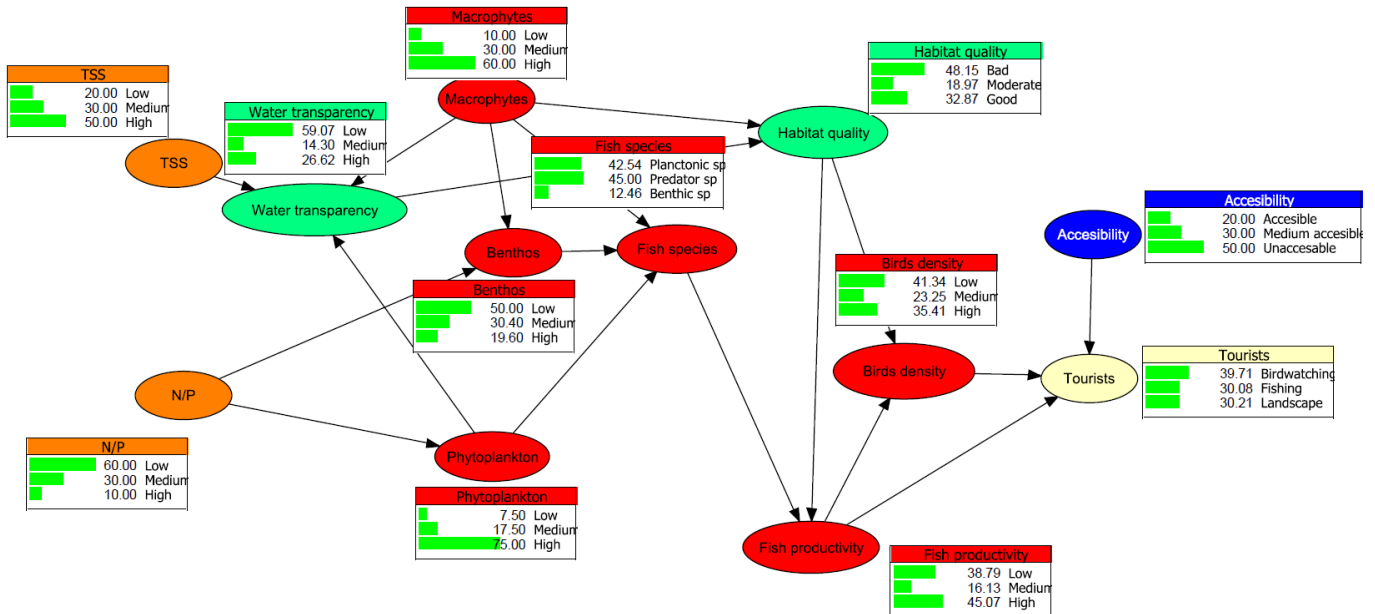


Figure 6.4-2: Bayesian Belief Network developed for assessing the trade-offs between provisioning services and the tourism and recreation in the Danube Delta. The nodes are grouped and coloured based on the types of variables they represent using a Drivers, Pressures (orange), State (green), Impact (red), Response (blue) approach model. Total suspended solids (TSS), phytoplankton (represented by chlorophyll-a) and macrophyte cover are both remote sensing and in-situ data. The outputs of the network are the tourists’ interests, bird densities and fish productivity. Arrows represent causalities, between different ecosystem properties and ES.

6.4.2 Quantifying the network (CPTs)

The probability tables were derived after agreement between experts and compared with literature.

Some of variables (e.g. concentration of chlorophyll a, nitrogen, phosphorus, fish and birds abundance etc.) were estimated both from historical data as well as through in-situ monitoring activities using exploratory statistics to determine the range of the variables and the relations among them.

6.4.3 Spatial application

The BBN is planned to be run spatially at full extent of the Danube Delta protected area, using an ecosystem distribution map, interpolated layers of variables measured in-situ (e.g. chlorophyll a, total suspended solids, fish and birds abundance, number of tourists) and remote sensing data describing the water quality (e.g. water chlorophyll content, water suspended solids, land surface temperature, and hydroperiod). The output resolution would be about 1 square km.



6.4.4 Description of the network nodes

Table 6.4-1: Description of nodes in the Danube Delta network, their states, and methods use to quantify their conditional probabilities.

	Description	States	CPT method
Nodes representing the pressure			
TSS	Total Suspended solids	3 states: 0 - 50	Expert knowledge (three-point estimation)
N/P	Nitrogen/ Phosphorous ratio (inverse relationship)	3 states: 10 - 60	Expert knowledge (three-point estimation)
Nodes representing the state			
Water transparency	Water transparency	3 states: 7.5 - 75	Expert knowledge (three-point estimation)
Habitat quality	Habitat quality	3 states:	Expert knowledge (three-point estimation)
Nodes representing the impact			
Macrophytes	Macrophytes	3 states: 10 – 60	Expert knowledge (three-point estimation)
Benthos	Benthos	3 states: 10 – 60	Expert knowledge (three-point estimation)
Phytoplankton	Phytoplankton	3 states: 50-19.6	Expert knowledge (three-point estimation)
Fish species	Fish species	3 states	Expert knowledge (three-point estimation)
Bird density	Bird density	3 states	Expert knowledge (three-point estimation)
Fish productivity	Fish productivity	3 states	Expert knowledge (three-point estimation)
Nodes representing the response			
Accessibility	Accessibility	3 states:20-50	Expert knowledge (three-point estimation)
Separate nodes			
Tourism	Tourism	3 states:	Expert knowledge (three-point estimation)

6.5 A Trade-off Analysis of Ecosystem Services and Dynamic BNs in the Dutch Wadden Sea, the Netherlands

Authors: Sonja Wanke, Alex Ziembra, Ghada el Serafy (Deltares)

Evaluating interactive effects of global factors with local pressures is one of the major challenges for future management of the Wadden Sea. Global pressures like sea level and temperature rise, and invasive species act on a large scale. For future management strategies it is vital to understand effects of global drivers and their influences on a regional level (van Beusekom, J. E., Buschmann, C. & Reise 2012). In combination with local anthropogenic activities, global pressures can have significant effects on the ecosystem. When modelling the marine environment and its ecosystem services, it is therefore vital to include global and regional pressures such as climate change (global pressures) and eutrophication, fisheries and pollution (regional pressures) that may frequently occur as non-linear interactions (Petersen et al. 2017). The following research was divided into two separate studies to limit its complexity. The first part dealt with a trade-off analysis between two different ecosystem services developing a rather large Bayesian Network. Whereas the second part developed a much smaller Dynamic Bayesian Network showing cumulative and cascading effect on the ecosystem and its services in the Dutch Wadden Sea.

6.5.1 Developing the Bayesian Network

In order to get a better qualitative understanding of the system, the risks for the system were identified with a DPSIR. Information was gathered from literature. The DWS is a complex system that has various *driving forces*. Anthropogenic drivers dominate the system over natural drivers. The most influential anthropogenic drivers are commercial fishery, water pollution, maintenance and other activities. Climate change is a global driver and the low lying coastal land is the result of tectonic movement. While both of these factors are not manageable, their impacts can be minimized.

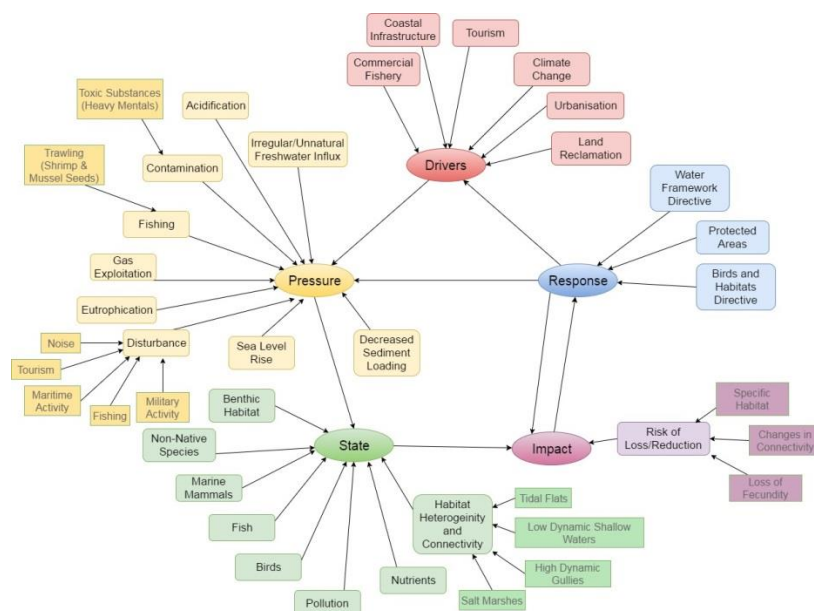


Figure 6.5-1: DPSIR Model of the Dutch Wadden Sea

Once the DPSIR framework and the possible trade-off analysis were done, a generic conceptual model of the Dutch Wadden Sea could be prepared. The aim of this exercise was to get a better understanding of the relationship

between the two chosen ESs. The red squares represent the drivers and therefore the starting point. The arcs between different types of squares depict the general direction of the cause-effect relationship within the system. The model includes different kinds of topics that affect fish populations (blue), bird watching (green) or both alike (yellow). The conceptual model is a fairly holistic view of the DWS as it includes factors that cover various time and spatial scales within the DWS. In order to build a BBN from the conceptual model, it needs further simplification, even though it already describes a simplified world. However, it is important to also be aware of the underlying causal relationships and the system complexity when building a BBN. Furthermore, making a conceptual model helps to identify which variables are impacted by which variables, hence illustrating the relevance of some factors such as ‘Forging Habitats’ over others (e.g. ‘Grazing’). All in all, this model helped to translate the immense amount of information identified above into more tangible variables.

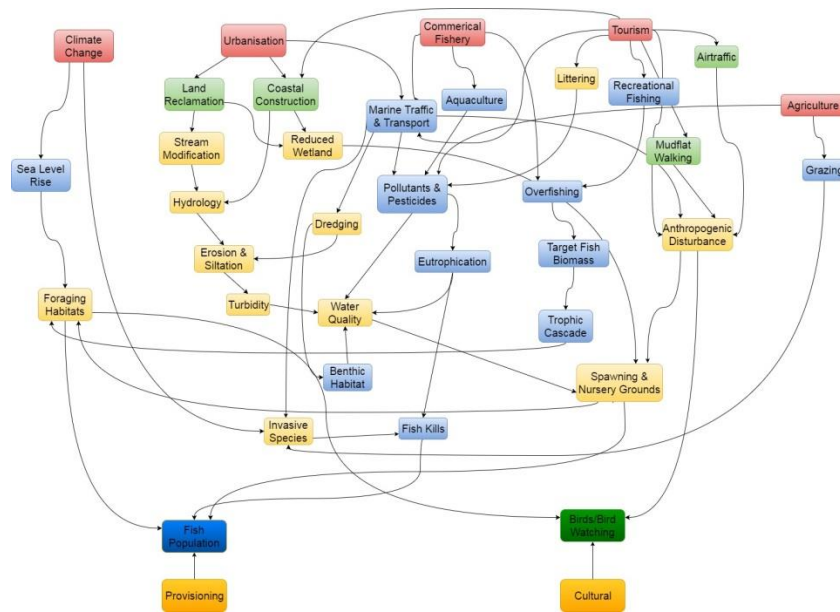


Figure 6.5-2: Conceptual Model of the Dutch Wadden Sea

6.5.2 Quantifying the network (CPTs)

Different types of data (e.g. expert knowledge, in-situ measurements) were integrated for the purpose of these two studies.

6.5.2.1 Data analysis

For any used data set, preparation is required as *Netica* learns data in a specific way. The Data preparation as filtering, formatting and validating for this research is carried out with *Excel*; other studies may also use *R* or *Python*. Some data sets consisted of a great amount of cases (more than 1 million) so in order to prepare this data the Add-In *Power Pivot* was used. Another Add-In that was helpful for the data preparation is *Power Query*, as it simplifies data fusion. The generic Framework for Data Analysis is as follows:

- Data inventory and collection
- Data cleaning and reduction
- Data Pre-processing and analysis



- Data Discretization
- Final file adjustment and export

6.5.2.2 Expert knowledge

Expert elicitation can be used to assist decision making before all necessary data is available and should not be used as a substitute (Pollino & Henderson, 2010). It is a rather demanding task for experts to assign a value of probability parameters as the human reasoning is rarely based on such probabilities (Kjærulff & Madsen, 2013). When using expert knowledge it is vital to document, defend and validate the model structure's basis and its conditional probabilities (McCann et al. 2006). Several methods exist for probability elicitation: iterative processes (combining the experts expressed knowledge); probability scales containing verbal and/or quantitative anchors; estimates of probability based on the expert's range of belief; and usage of frequencies. Participants that have a mathematical background provide a more accurate estimation on their beliefs based on the expert's range of belief (Constantinou et al. 2016).

The expert knowledge was implemented into the model with the following steps:

1. Link the expert variable to the child and (optional) parent variables.
2. Using expert judgement parameterize the expert variable's CPT
3. Each child variables' CPT is parameterized with the expert knowledge
4. r_{im} -Equation is then used to learn the probability distribution for all to each state m of the child variables (Constantinou et al. 2016).

The scalability of this method can be determined by various factors. The first one is the *type of model* as it is only useful for models that are incorporating expert judgement and features of data driven models. Additionally, the method is independent from the *size of the model*. Though, it is of importance how the expert variables are introduced. It is less challenging to link two data driven variables with an expert variables when the network consists of thousands of variables. On the contrary is more challenging to link three data –driven variables with an expert variable when the network has only three variables.

6.5.3 Spatial application

The trade-off analysis BN has a spatial node (habitat). At this stage the CPT was filled in by estimating the sizes of each habitat. In the future more accurate land cover and remote sensing data will be used.

6.5.4 Dynamics

The construction of a DBN model is not a linear process but rather an ongoing cycle in which new information from field tests, literature and other sources can be included once they become available. A conceptual model of the system should be created prior to building a BN which synthesizes existing knowledge. It is possible to build various conceptual models when showing different level of detail, perspectives or scales. This stage provides a visualized summary of drivers and their direct and indirect relations (link) to other variables and outputs. The conceptual model can help expose weaknesses, therefore (Jakeman et al. 2006) advised to always undergo the conceptualisation step. Experts are able to review the conceptual model after the model has been built. Their

feedback may also help to recognize the most important variables, processes and possible errors can be corrected (Chen & Pollino, 2012).

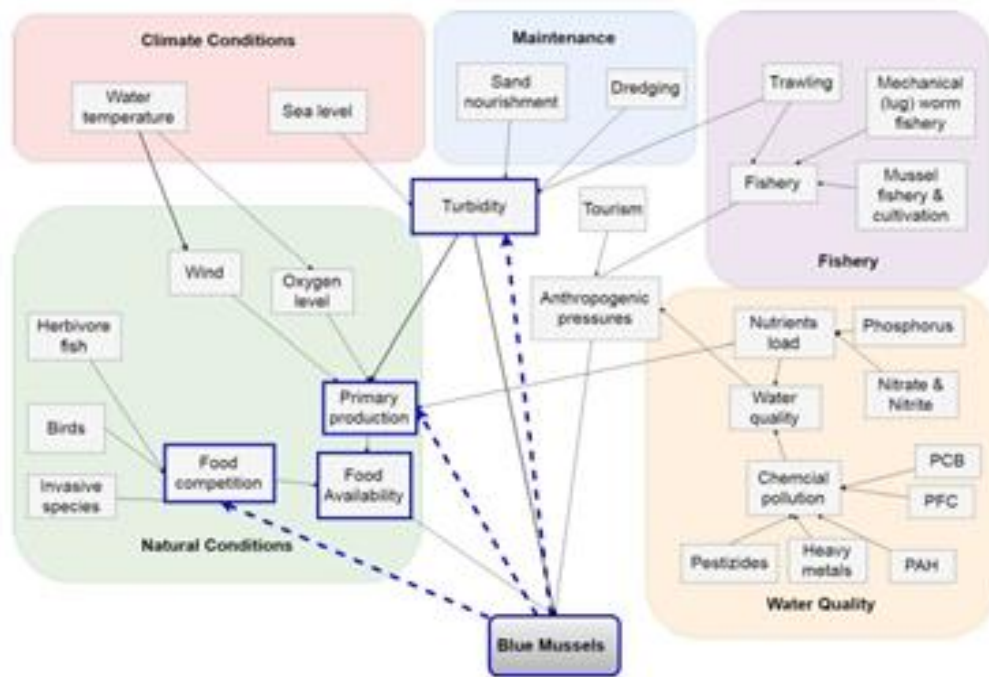


Figure 6.5-3: Conceptual Dynamic Bayesian Network

After this step, the physical variables that are to be included are chosen and it is decided how detailed the model is supposed to be. When defining the variables it is important to do it in a way that all users are able to understand the represented variables (Kragt 2009). Kragt (2009) and Marcot et al. (2006) recommended to keep the number of parent variables to three or fewer in order to bound the CPT size. For the same reason, their states should be maximum five (Marcot, et al., 2006). These recommendations keep the CPT small enough to be tractable and understandable. A large amount of intermediate variables, also known as latent variables, will contain uncertainty propagation.

6.5.5 Validation and sensitivity analysis

Both networks were validated by experts multiple times. As a next step a sensitivity analysis using Netica is planned to be performed.

6.5.6 Description of the network nodes

6.5.6.1 Using Bayesian Networks to Capture Trade-offs Between Ecosystem Services

During the first phase, good background knowledge about the study area and its resulting ESs were gained. Additionally, information on the directional correlation between mussel dynamics and bird populations was acquired. On the basis of that a literature based BN was constructed. The whole network was comprised of more than 40 nodes which had an effect on each other in one way or another. Each node was either trained with data or expert knowledge. To train the network, measurements from the last 10 years were taken. Monthly and weekly

Averages were determined during the data analysis to see whether there are big differences between these two distributions.

The proposed literature based BN conceptual model structure was improved and enhanced and turned it into a fully functional BN model. This was done by consulting with experts from Deltares, the Netherlands. The first round of expert elicitations focused on a discussion of the proposed literature based BN model as well as its purpose. Furthermore, the suitability of the chosen nodes and links was discussed. This was done via face-to-face meetings. These meetings led to the identification of superfluous nodes which were defined during the literature review, but were deemed inappropriate in the context of the Dutch Wadden Sea and therefore exchanged, deleted, or redefined.

Different test and experiments were conducted once the network was trained to see how certain nodes would affect other nodes. For example, not only parent nodes can be filtered. The BN is able to show the probabilities of parent and even “grandparent” nodes when a child node is filtered. Looking at a snippet of the whole BN, when the Commercial Fishing node is set to a very low state, it can be seen that this implies a certain state in the parent nodes (Mussel Stock) and even “grandparent” nodes (NPP, O₂, and Water Temperature). In this case the model assumes a low mussel stock when Commercial Fishing is decreased. In case the fishing activity is set to very high, the fish stock increases. It is worthwhile to mention that this assumption was not entered into the model at any time but is an effect of the given and trained data.

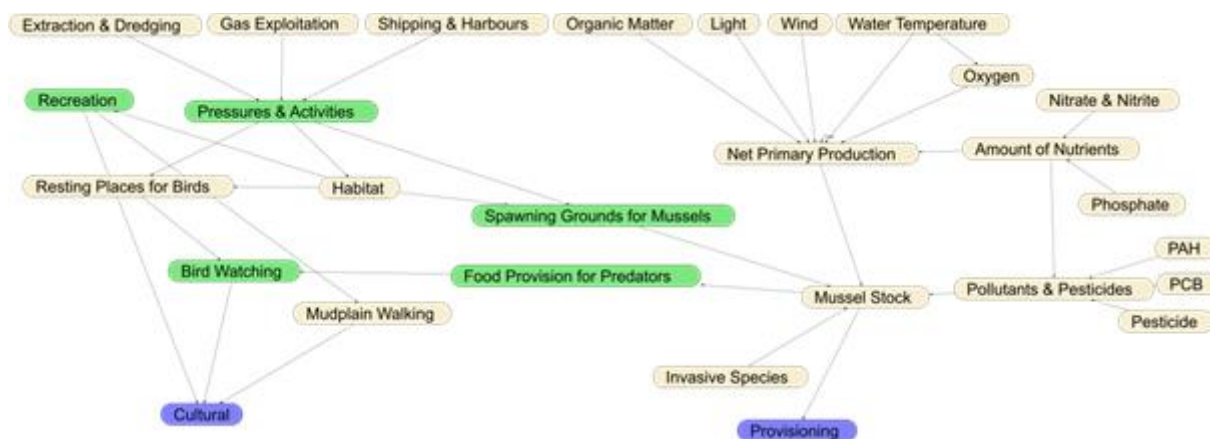


Figure 6.5-4: Complete BN Structure

Table 6.5-1: Description of nodes in the network, their states, and methods use to quantify their conditional probabilities.

Node	Description	States	CPT method
Data Trained Nodes			



Oxygen level of the water	States were determined by dissolved oxygen (DO) as it acts as physio and chemical supporting element in (salt) water	4 states: - Low - Medium - High - Lethal
Wind	Mixes the upper layers of a water body by increasing tides and wave heights (Reise, et al., 2010)	9 states according to the Beaufort scale
Water temperature	Is vital for many biological processes such as primary production which affects the algal growth rate.	5 states: - Extreme low - Low - Medium - High - Extreme high
Organic matter	It is matter composed of organic compounds which come from dead organisms and their waste (detritus).	3 states: - Low - Medium - High
Light	A fundamental factor in the ecology of aquatic ecosystems is underwater light as the quantity of light is crucial for photosynthesis for both benthic and pelagic primary producer (Obrador and Pretus 2008)	3 states: - Low - Medium - High
Net primary production	Contain chlorophyll-a, which can therefore be used as an indicator for PP. This node is important as it represents the lowest trophic level and with that the foundation of a food web (McDonald et al. 2016)	3 states: - Low - Medium - High
Phosphate	Is required for algal accumulation and growth, such as phytoplankton, in the water column alongside other nutrients (McDonald, et al., 2016) and if its concentration is too high it can cause eutrophication (Reise et al. 2010)	3 states: - Low - Medium - High
Nitrate	It is vital for algal accumulation and growth (McDonald, et al., 2016)	3 states: - Low - Medium - High
PCB	These chemicals can cause cancer and problems with the nervous system in humans and can be detected in fish and food products.	3 states: - Low - Medium - High
PAH	Their toxicity is moderate to marine organisms but can produce carcinogenic metabolites in organisms (Cole et al. 1999)	3 states: - Low - Medium - High



Amount Nutrients	of	Due to anthropogenic activities such as agriculture, industry and domestic effluent cause an increase in nutrient load in the ecosystems	3 states: - Low - Medium - High	Summarising node
Pesticides		Are chemicals that are used as insecticides, fungicides, herbicides and rodenticides (Eddleston 2016) available for usage in the 1940.	2 states: - Below norm - Above norm	
Pollutants Pesticides	&	Summarizes the concentration pesticides and other pollutants such as heavy metals, PCB, PAH and PFC.	2 states: - Good - Bad	Summarising node
Gas Exploitation		Gas extraction takes place in the WS.	3 states: - Low - Medium - High	
Shipping Harbours	&	Even though numbers are declining, accidents at sea and the illegal disposal of harmful substances still constitute a potential danger to the environment in the WS area.	3 states: - Low - Medium - High	
Habitat		Depending on the habitat different species, ecosystem services and activities can be found.	5 states: - Bad Tidal flat - Salt marshes - Near shore North Sea - High dynamic gullies - Low dynamic gullies - Dunes	
Mudplain Walking		One of the main tourist activities in the WS.	2 states: - Yes - No	
Invasive species		May pose a risk on the natural existing bivalve species as they have a higher growth rate and recruitment success, and the absence of natural predators (Nehls et al. 2009)	3 states: - Low - Medium - High	
Resting places for birds		Are attracted to the WS because it is an area with high food availability such as bivalve species, worms (Eriksson et al. 2010) or fish.	3 states: - Low - Medium - High	
Expert Trained Nodes				
Recreation		Every year, tourist visits the WS and DWS which results in increased disturbance and ongoing development of the static coastline (CWSS 2012).	2 states: - Attractive - Not attractive	Experts knowledge (3-point expert elicitation)



Food provision for predators	It is also influenced by itself and the available blue mussel the predator may feed on.	2 states: - Low - High	Experts knowledge (3-point expert elicitation)
Spawning grounds for mussels	Vital for mussels.	2 states: - Sufficient - Insufficient	Experts knowledge (3-point expert elicitation)
Pressures & Activities	Is the sum of human activities such as tourism, commercial fishery, and water pollution.	3 states: - Low - Medium - High	Experts knowledge (3-point expert elicitation)
Bird watching	It is one of the main tourist attractions in the WS.	2 states: - Possible - Not possible	Experts knowledge (3-point expert elicitation)

6.5.6.2 Using Dynamic Bayesian Networks to Capture Cumulative Effect on Ecosystem Services

After carrying out expert surveys some changes had been made from the original conceptual model to the BN created as the first step in the design phase (see Figure 6.5-5). It was pointed out that the recruitment success and standing stock is highly dependent on predator (such as birds, fish, shrimp and sea stars). Another change was made for nutrient load which is highly connected to primary production and has a positive effect on mussel biomass. On the other hand, chemical pollution has rather a negative effect. Therefore, nutrient load is not connected to water quality anymore. As this node has only one input left (chemical pollution) it could be removed from the network in order to simplify it.

When defining the variables for this BN, data sources were taken into account and whether a variable is static or dynamic. There follows, the CPT's for the expert elicitations (turquoise). The blue coloured variables CPT's are learned with in-situ data apart from the *Nutrient Load* and *Chemical Pollution* variables which are summarizing variables (red). Most of the temporary (dynamic) variables CPT's are also trained with in-situ data (pink): *Turbidity*, *Primary Production*, *Predators*, and *Blue Mussel* (Blue), apart from *Food Availability* (Purple) which is trained with the help of Expert Knowledge.

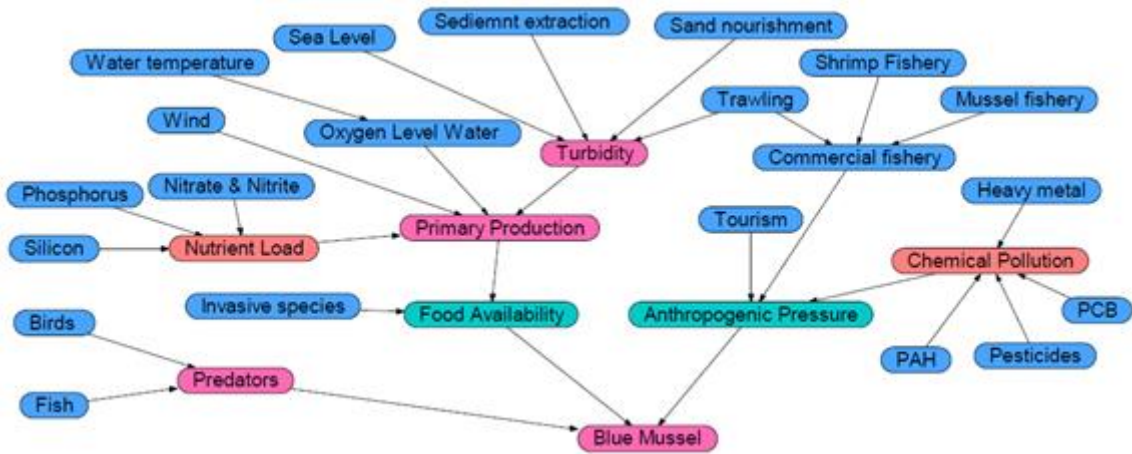


Figure 6.5-5: BN Various Anthropogenic Pressures and Natural Conditions Affecting Blue Mussel Abundance

Figure 6.5-5 shows the developed DBN of this study. The blue coloured variables are primarily derived by in-situ data. The pink coloured variables represent the temporal variables. The variables of the first time-slice are: *Turbidity [0]*, *Primary Production [0]*, *Food Availability [0]*, *Predators [0]*, and *Blue Mussel [0]*. The second time slice contains the variables *Turbidity [1]*, *Primary Production [1]*, *Food Availability [1]*, *Predators [1]*, and *Blue Mussel [1]*. Expert variables are the *Food Availability* in both Time slices and *Anthropogenic Pressure*.

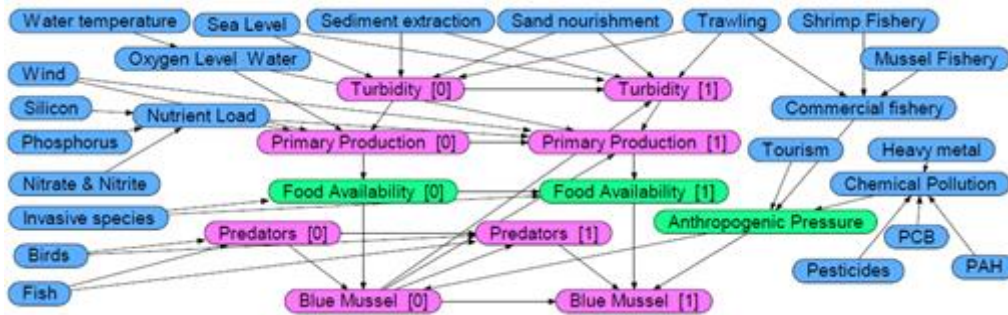


Figure 6.5-6: DBN Various Anthropogenic Pressures and Natural Conditions Affecting Blue Mussel Abundance - Post Time Expansion

Table 6.5-2: Description of nodes in the Wadden Sea network, their states, and methods use to quantify their conditional probabilities (showing only the nodes that were not used in the static BN)

Node	Description	States	CPT method
Static nodes			
Sea level	Indicator for this node is the mean daily sea level	3 states: - Low - Medium - High	



Fish	Herbivory species are representing the second trophic state and feed on primary producers. Higher trophic fish species are predators to bivalves or species that compete for the same food sources.	3 states: - Low - Medium - High	
Sand nourishment	Non-native sand is added on the beach and may change the beach's morphology and could lead to local negative effect on macroinvertebrate communities, also leading to a higher amount of suspended matter (Leewis et al. 2012)	3 states: - Low - Medium - High	
Commercial Fishery	The fishing quota was used as an indicator for this node which is determined by the Total Allowable Catches (TACs) that are annually set for the fish stock by the Council of fisheries ministers	2 states: - Higher - Lower	
Mussel fishery & cultivation	In the DWS, some unstable and young beds outside the protected areas were restricted to experimental fishery.	3 states: - Increased - Decreased	
Shrimp fishery	Can have a negative effect on mussel as they may eat their larvae that may have a positive effect on blue mussel.	2 states: - Higher - Lower	
Trawling	Its negative effects on benthic habitat are dependent on gear configurations, sea floor characteristics, depth, and sensitivity of the species that form the habitat. The response of the effected species and the recovery of the habitat vary among taxa (Malecha and Heifetz 2017).	2 states: - Higher - Lower	
Silicon	Can co-regulate trophic shifts along other nutrients such as phosphorus, nitrogen and carbon (McDonald, et al., 2016), and contributes to stimulate algal growth and deoxygenation of the water column and sediment (Cole, et al., 1999)	3 states: - Low - Medium - High	
Heavy metal	Are carcinogenic and toxic to humans and animals.	2 states: - High above legal concentration - below legal concentration	
Chemical Pollution	Summarizes the concentration pesticides and other pollutants such as heavy metals, PCB, PAH and PFC.	2 states: - Good - Bad	Summarizing node

Dynamic nodes



Turbidity	A fundamental factor in the ecology of aquatic ecosystems is underwater light as the quantity of light is crucial for photosynthesis for both benthic and pelagic primary producer (Obrador and Pretus 2008).	3 states: - Low - Medium - High	
Predators	It is also influenced by itself and the available blue mussel the predator may feed on.	2 states: - Low - High	summarizing node
Blue Mussel	Can be used to represent overall effects on the ecosystem	2 states: - Low - High	

6.6 Whale watching (Pelagos)

Authors: Kevin Flückiger, Ana Stritih (ETH Zürich)

The Pelagos Sanctuary is a marine protected area (MPA) in the Mediterranean Sea aimed to protect all marine mammals. Due to a combination of climatic, oceanographic and physiographic factors, the area has good conditions for the feeding and breeding habitat of several whale and dolphin species (Notarbartolo di Sciara and Reeves 2006). At the same time, it is surrounded by well-developed regions with high economic activity. For tourists and locals, whale watching has become an increasingly popular activity, and is therefore an important cultural ecosystem service in the MPA. However, cetaceans in the area are threatened by pollution, ship strikes, and noise. Disturbance due to whale watching is also thought to be harmful to whales, but data on the activity and its impacts are scarce. A BN was developed to map the value of whale watching by combining data on cetaceans with expert knowledge on whale watching activities in the Pelagos Sanctuary.

6.6.1 Developing the Bayesian Network

A draft network was initially constructed through literature review, particularly regarding threats to cetaceans. Since literature on whale-watching behaviour and value of cetacean sightings is lacking, this part of the network was developed based on a workshop with stakeholders (local researchers, conservationists, and whale watching companies). Then, the network was developed further in an iterative process of discussion with experts, collecting data, and testing.

The network contains nodes on cetacean presence (sightings and potential feeding grounds), threats to cetaceans, whale-watching behaviour, and the social value of cetacean sightings.

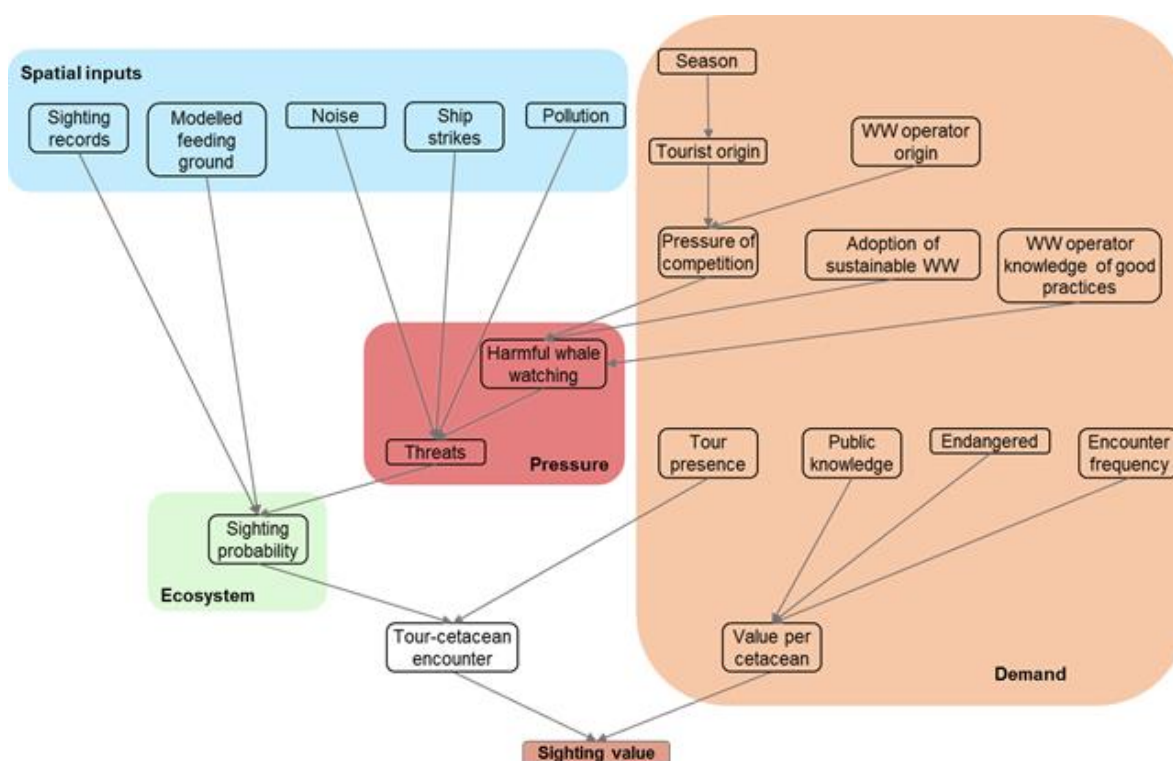


Figure 6.6-1: Bayesian Network developed to model the ES of whale watching

6.6.2 Quantifying the networks (CPTs)

Expert interviews were conducted to populate nodes describing the behaviour of whale watching companies, their harmfulness to cetaceans, and the impacts of other threats on cetacean wellbeing. Furthermore, the social value of cetacean sightings (dependant on the level of public knowledge, perceived cetacean vulnerability, and frequency of encounters) was estimated based on expert knowledge.

6.6.3 Spatial application

The BN was applied on a 4x4 km raster of the Pelagos MPE, using gBay. The spatial inputs to the BN are data on cetacean presence (reported sightings and modelled suitability of feeding ground), spatial data threats to cetaceans (shipping, noise), and information on whale-watching tour locations that was collected at the stakeholder workshop.

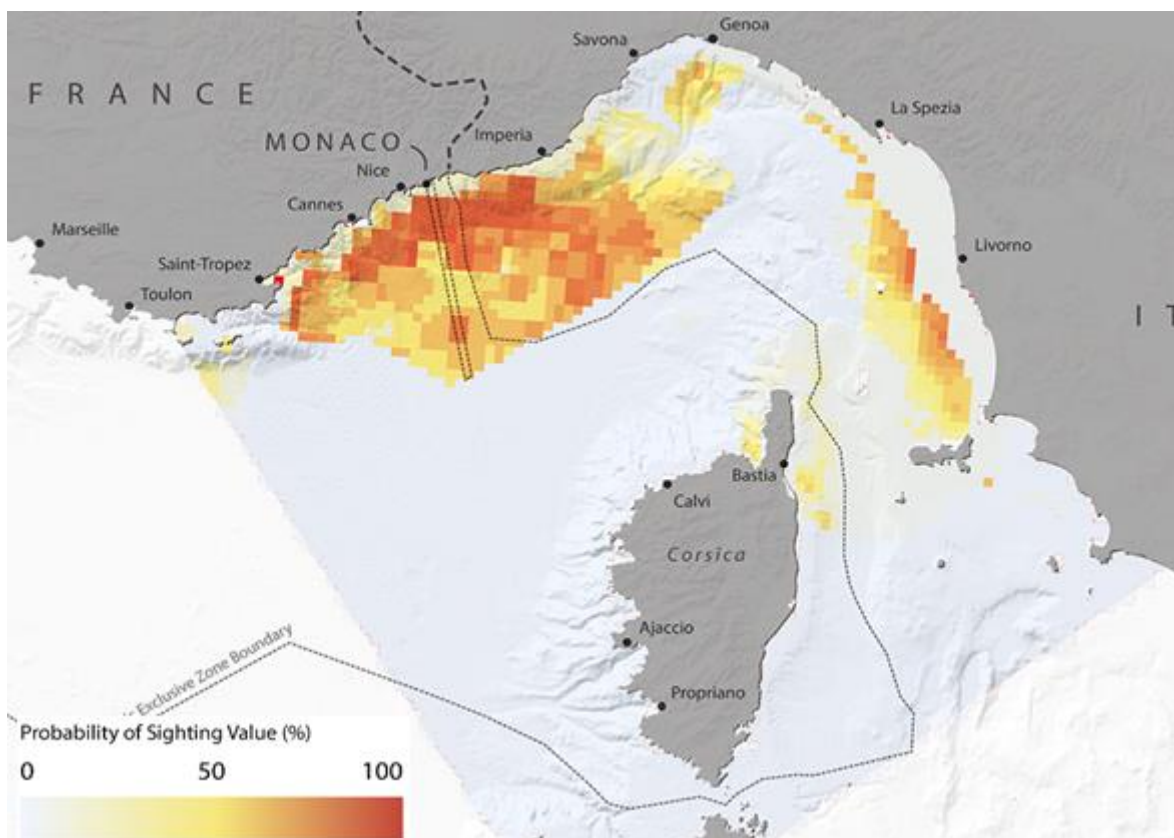


Figure 6.6-2: A map of the "Sighting Value" of whale watching in the Pelagos Sanctuary, calculated using the BN.

6.6.4 Validation and sensitivity analysis

A sensitivity analysis of the node "Sighting value" was performed using the Netica function "Sensitivity to findings". Based on the sensitivity analysis, the most important threat to cetaceans is pollution. Whale-watching operators' knowledge is an important factor in reducing the negative impacts of whale watching.



6.6.5 Description of the network nodes

Table 6.6-1: Description of nodes in the avalanche protection Bayesian Network, their states, and methods use to quantify their conditional probabilities.

Node	Description	States	CPT method
Spatial inputs			
Possible ground	feeding Probability derived from habitat suitability model by (Druon et al., 2012)	Boolean	
Sighting records	Combination of sighting records from various sources	Boolean	
Harmful noise	Noise from harbours, military exercises, offshore drilling, seismic surveys, mapped by ACCOBAMS.	Boolean	
Harmful (shipping)	pollution Pollution from maritime traffic, mapped in the Med-IAMER project	Boolean	
High risk (ship strikes)	unnatural death Risk of ship strikes (Vaes and Druon, 2013).	Boolean	
Pressure			
Harmful watching	whale Harmful actions by whale-watchers, e.g. approaching the cetaceans to closely.	Often Rare	Expert estimate
Threats to wellbeing	cetacean Combined probability of threats.	Boolean	Combination of threats, using weights defined by experts
Demand			
Tour presence	Spatial data on tour locations, collected through participatory mapping at stakeholder workshop.	Boolean	
WW operator knowledge about good behaviour	Personal knowledge of the boat operators and management about non-harmful behaviour.	Good Bad	Expert estimate
Adaption of Sustainable WW label	of The code of conduct is defined by the Pelagos sanctuary board of administration.	Labelled No label	Expert estimate
Season	The season affects the visitors (more locals off-season)	Offseason Season	



Tourist origin	Local visitors have higher expectations.	Local Non-local	Expert estimate
WW operator origin	Higher pressure of competition among French WW companies.	French Italian	Expert estimate
Pressure of competition	The feeling of whale watching operators to be under pressure to deliver a better experience to customers than other operators, caused by the business environment. High pressure reduces the likelihood to follow any rules.	High Low	Expert estimate
Public knowledge trend	The amount of knowledge of the public about cetaceans.	High Not informed	Expert estimate
Perceived cetacean vulnerability	The perception of the cetacean vulnerability in the Pelagos sanctuary by different stakeholder groups.	Endangered Not endangered	Expert estimate
Encounter frequency	An encounter with cetaceans is more highly valued if such encounters are rare.	Rare Frequent	Expert estimate
Value per cetacean	Social value of seeing a cetacean.	High Medium Low	Expert estimate
Potential ES provision			
Sighting probability	Probability of cetacean presence, reduced in case of threats	Boolean	Expert estimate
Ecosystem service			
Tour-cetacean encounter	Tour and cetacean in the same location at the same time.	Boolean	Combination of tour presence and sighting probability
Sighting value	Value of cetacean sighting.	High Medium Low None	Combination of encounter and value per cetacean.



7. References

- Ames, D. P., B. T. Neilson, D. K. Stevens, and U. Lall. 2005. Using Bayesian networks to model watershed management decisions : an East Canyon Creek case study. *Journal of Hydroinformatics* 7(4):267–282.
- Bebi, P., F. Kienast, and W. Schönenberger. 2001. Assessing structures in mountain forests as a basis for investigating the forests' dynamics and protective function. *Forest Ecology and Management* 145(1–2):3–14.
- Bebi, P., D. Kulakowski, and C. Rixen. 2009. Snow avalanche disturbances in forest ecosystems-State of research and implications for management. *Forest Ecology and Management* 257(9):1883–1892.
- van Beusekom, J. E., Buschmann, C. & Reise, K. 2012. Wadden Sea tidal basins and the mediating role of the North Sea in ecological processes: Scaling up of management?
- Bromley, J. 2005. *Guidelines for the use of Bayesian networks as a participatory tool for Water Resource Management*. Page *Management of the Environment and Resources using Integrated Techniques*. Centre for Ecology and Hydrology, Wallingford.
- Cain, J. 2001. *Planning improvements in natural resources management: Guidelines for using Bayesian networks to support the planning and management of development programmes in the water sector and beyond*. Centre for Ecology and Hydrology, Wallingford.
- Castelletti, A., and R. Soncini-Sessa. 2007. Bayesian Networks and participatory modelling in water resource management. *Environmental Modelling and Software* 22(8):1075–1088.
- Celio, E., and A. Grêt-Regamey. 2016. Understanding farmers' influence on land-use change using a participatory Bayesian network approach in a pre-Alpine region in Switzerland. *Journal of Environmental Planning and Management* 59(11):2079–2101.
- Celio, E., T. Koellner, and A. Grêt-Regamey. 2014. Modeling land use decisions with Bayesian networks: Spatially explicit analysis of driving forces on land use change. *Environmental Modelling and Software* 52:222–233.
- Chen, S. H., and C. a. Pollino. 2012. Good practice in Bayesian network modelling. *Environmental Modelling & Software* 37:134–145.
- Christen, M., J. Kowalski, and P. Bartelt. 2010. RAMMS: Numerical simulation of dense snow avalanches in three-dimensional terrain. *Cold Regions Science and Technology* 63(1–2):1–14.
- Cole, S., I. Codling, W. Parr, and T. Zabel. 1999. Guidelines for managing water quality impacts within UK European marine sites. Swindon: UK Marine SAC Project.
- Constantinou, A. C., N. Fenton, and M. Neil. 2016. Integrating expert knowledge with data in Bayesian networks: Preserving data-driven expectations when the expert variables remain unobserved. *Expert Systems with Applications* 56:197–208.
- Cripps, E., A. O Hagan, T. Quaife, and C. Anderson. 2009. *Modelling uncertainty in satellite-derived land cover maps*. Page *Research Report No. 573/08*. Department of Probability and Statistics, University of Sheffield.
- CWSS. 2012. The Wadden Sea, Germany and Netherlands (N1314) - Extension Denmark and Germany.
- Das, B. 2004. Generating Conditional Probabilities for Bayesian Networks: Easing the Knowledge Acquisition Problem. *CoRR*:1–24.



- Díaz, G. I., L. Nahuelhual, C. Echeverría, and S. Marín. 2011. Drivers of land abandonment in Southern Chile and implications for landscape planning. *Landscape and Urban Planning* 99(3–4):207–217.
- Eddleston, M. 2016. Pesticides. *MEDICINE* 44(3).
- Eriksson, B. K., T. van der Heide, J. van de Koppel, T. Piersma, H. W. van der Veer, and H. Olf. 2010. Major changes in the ecology of the wadden sea: Human impacts, ecosystem engineering and sediment dynamics. *Ecosystems* 13(5):752–764.
- Feistl, T., P. Bebi, M. Teich, Y. Bühler, M. Christen, K. Thuro, and P. Bartelt. 2014. Observations and modeling of the braking effect of forests on small and medium avalanches. *Journal of Glaciology* 60(219):124–138.
- Grêt-Regamey, A., S. H. Brunner, J. Altwegg, and P. Bebi. 2013. Facing uncertainty in ecosystem services-based resource management. *Journal of Environmental Management* 127:145–154.
- Grêt-Regamey, A., and D. Straub. 2006. Spatially explicit avalanche risk assessment linking Bayesian networks to a GIS. *Natural Hazards and Earth System Science* 6(6):911–926.
- Hamilton, S. H., C. A. Pollino, and A. J. Jakeman. 2015. Habitat suitability modelling of rare species using Bayesian networks: Model evaluation under limited data. *Ecological Modelling* 299:64–78.
- Jakeman, A. J., R. A. Letcher, and J. P. Norton. 2006. Ten iterative steps in development and evaluation of environmental models. *Environmental Modelling and Software* 21(5):602–614.
- Kjaerulff, U., and A. Madsen. 2013. *Bayesian Networks and Influence Diagrams*. Springer.
- Kleemann, J., E. Celio, B. K. Nyarko, M. Jimenez-Martinez, and C. Fürst. 2017. Assessing the risk of seasonal food insecurity with an expert-based Bayesian Belief Network approach in northern Ghana, West Africa. *Ecological Complexity* 32:53–73.
- Kragt, M. E. 2009. A beginners guide to Bayesian network modelling for integrated catchment management. *Landscape Logic Technical Report* 9.
- Lamarque, P., A. Artaux, C. Barnaud, L. Dobremez, B. Nettièr, and S. Lavorel. 2013. Taking into account farmers' decision making to map fine-scale land management adaptation to climate and socio-economic scenarios. *Landscape and Urban Planning* 119:147–157.
- Leewis, L., P. M. van Bodegom, J. Rozema, and G. M. Janssen. 2012. Does beach nourishment have long-term effects on intertidal macroinvertebrate species abundance? *Estuarine, Coastal and Shelf Science* 113:172–181.
- Malecha, P., and J. Heifetz. 2017. Long-term effects of bottom trawling on large sponges in the Gulf of Alaska. *Continental Shelf Research* 150(August):18–26.
- Marcot, B. G. 2012. Metrics for evaluating performance and uncertainty of Bayesian network models. *Ecological Modelling* 230:50–62.
- Marcot, B. G., R. S. Holthausen, M. G. Raphael, M. M. Rowland, and M. J. Wisdom. 2001. Using Bayesian belief networks to evaluate fish and wildlife population viability under land management alternatives from an environmental impact statement. *Forest Ecology and Management* 153(1–3):29–42.
- Marcot, B. G., J. D. Steventon, G. D. Sutherland, and R. K. McCann. 2006. Guidelines for developing and updating Bayesian belief networks applied to ecological modeling and conservation. *Canadian Journal of Forest Research* 36(12):3063–3074.



- McCann, R. K., B. G. Marcot, and R. Ellis. 2006. Bayesian belief networks: applications in ecology and natural resource management. *Canadian Journal of Forest Research* 36(12):3053–3062.
- McDonald, K. S., M. Tighe, and D. S. Ryder. 2016. An ecological risk assessment for managing and predicting trophic shifts in estuarine ecosystems using a Bayesian network. *Environmental Modelling & Software* 85:202–216.
- Mendez, P. F., N. Isendahl, J. M. Amezaga, and L. Santamaria. 2012. Facilitating transitional processes in rigid institutional regimes for water management and wetland conservation: experience from the Guadalquivir Estuary. *Ecology and Society* 17(1):26.
- Méndez, P. F., L. Santamaría, J. Amezaga, and G. Hearn. 2010. *Adaptive strategies for natural resources and ecosystems in Canada - Opportunities for implementation in Europe*.
- Merz, H. A., T. Schneider, and H. Bohnenblust. 1995. *Bewertung von technischen Risiken: Beiträge zur Strukturierung und zum Stand der Kenntnisse: Modelle zur Bewertung von Todesfallrisiken*. ETH, Zürich.
- Metcalf, S. J., and K. J. Wallace. 2013. Ranking biodiversity risk factors using expert groups - Treating linguistic uncertainty and documenting epistemic uncertainty. *Biological Conservation* 162:1–8.
- Nehls, G., S. Witte, H. Buttger, N. Dankers, J. Jansen, G. Millat, M. Herlyn, A. Markert, P. Sand Kristensen, M. Ruth, C. Buschbaum, and A. Wehrmann. 2009. Beds of blue mussels and Pacific oysters. Thematic Report No. 11. Pages 1–29 in H. Marencic and J. de Vlas, editors. *Quality Status Report 2009. WaddenSea Ecosystem No. 25*. Common Wadden Sea Secretariat, Trilateral Monitoring and Assessment Group, Wilhelmshaven, Germany.
- Newton, A. C., G. B. Stewart, A. Diaz, D. Golicher, and A. S. Pullin. 2007. Bayesian Belief Networks as a tool for evidence-based conservation management. *Journal for Nature Conservation* 15(2):144–160.
- Notarbartolo di Sciara, G., and R. Reeves. 2006. The status and distribution of Cetaceans in the Black sea and Mediterranean sea-Workshop report 5-7 March 2006. *The World Conservation Union IUCN*:137p.
- Obrador, B., and J. L. Pretus. 2008. Light regime and components of turbidity in a Mediterranean coastal lagoon. *Estuarine, Coastal and Shelf Science* 77(1):123–133.
- Pérez-Miñana, E. 2016. Improving ecosystem services modelling: Insights from a Bayesian network tools review. *Environmental Modelling and Software* 85:184–201.
- Petersen, M. E., M. Maar, J. Larsen, E. F. Moller, and P. J. Hansen. 2017. Trophic cascades of bottom-up and top-down forcing on nutrients and plankton in the Kattegat, evaluated by modelling. *Journal of Marine Systems* 169:25–39.
- Petrou, Z. I., V. Kosmidou, I. Manakos, T. Stathaki, M. Adamo, C. Tarantino, V. Tomaselli, P. Blonda, and M. Petrou. 2013. A rule-based classification methodology to handle uncertainty in habitat mapping employing evidential reasoning and fuzzy logic. *Pattern Recognition Letters* 48:24–33.
- Ramírez, F., I. Afán, L. S. Davis, and A. Chiaradia. 2017. Climate impacts on global hot spots of marine biodiversity. *Science Advances* 3(2):1–8.
- Regan, H. M., M. Colyvan, and M. A. Burgman. 2002. A taxonomy and treatment of uncertainty for ecology and conservation Biology. *Ecological Applications* 12(2):618–628.
- Reise, K., M. Baptist, P. Burbridge, N. Dankers, L. Fischer, B. Flemming, A. P. Oost, and C. Smit. 2010. The Wadden Sea - A Universally Outstanding Tidal Wetland: The Wadden Sea Quality Status Report Synthesis Report



2010. *Wadden Sea Ecosystem*(29):1–3.

Renwick, A., and C. Revoredo-Giha. 2008. Measuring Cross-Subsidisation of the Single Payment Scheme in England. *The Common Agricultural Policy after the Fischler Reform: National Implementations, Impact Assessment and the Agenda for Future Reforms*(January):1–16.

Salliou, N., C. Barnaud, A. Vialatte, and C. Monteil. 2017. A participatory Bayesian Belief Network approach to explore ambiguity among stakeholders about socio-ecological systems. *Environmental Modelling and Software* 96(June):199–209.

Speirs-Bridge, A., F. Fidler, M. McBride, L. Flander, G. Cumming, and M. Burgman. 2010. Reducing overconfidence in the interval judgments of experts. *Risk Analysis* 30(3):512–523.

Stelzenmüller, V., J. Lee, E. Garnacho, and S. I. Rogers. 2010. Assessment of a Bayesian Belief Network-GIS framework as a practical tool to support marine planning. *Marine Pollution Bulletin* 60(10):1743–1754.

Stritih, A., P. Bebi, and A. Grêt-Regamey. 2019. Quantifying uncertainties in EO-based ecosystem service assessments. *Environmental Modelling & Software* 111(January):300–310.

Uusitalo, L. 2007. Advantages and challenges of Bayesian networks in environmental modelling. *Ecological Modelling* 203(3–4):312–318.

Zadeh, L. A. 1965. Fuzzy sets. *Information and Control* 8(3):338–353.

Zamora, R., A. J. Pérez-Luque, F. J. Bonet, J. . Barea-Azcón, and R. Aspizua. 2016. *Global Change Impacts in Sierra Nevada: Challenges for Conservation*. Page *Global change impacts in Sierra Nevada: Challenges for conservation*.