Integrated network models for predicting ecological thresholds:

McDonald, K. S.; Turk, V.; Mozetič, P.; Tinta, T.; Malfatti, F.; Hannah, David; Krause, Stefan

DOI:
10.1016/j.envsoft.2017.01.017

Document Version
Peer reviewed version

Citation for published version (Harvard):
McDonald, KS, Turk, V, Mozetič, P, Tinta, T, Malfatti, F, Hannah, D & Krause, S 2017, 'Integrated network models for predicting ecological thresholds: Microbial – carbon interactions in coastal marine systems', *Environmental Modelling and Software*, vol. 91, pp. 156-167. https://doi.org/10.1016/j.envsoft.2017.01.017

Integrated network models for predicting ecological thresholds: microbial – carbon interactions in coastal marine systems

K.S. McDonald[AD], V. Turk[B], P. Mozetič[B], T. Tinta[B], F. Malfatti[C], D.M. Hannah[A], S. Krause[A]

[A]School of Geography, Earth and Environmental Sciences, University of Birmingham, Edgbastion, Birmingham, B15 2TT, UK.

[B] Marine Biology Station Piran, National Institute of Biology, Piran, 6330, Slovenia.

[C] OGS (Istituto Nazionale di Oceanografia e Geofisica Sperimentale - National Institute of Oceanography and Experimental Geophysics), Trieste, 34010, Italy.

[D] Corresponding author: k.mcdonald@bham.ac.uk

Abstract

This proof of concept study presents a Bayesian Network (BN) approach that integrates relevant biological and physical-chemical variables across spatial (two water layers) and temporal scales to identify the main contributing microbial mechanisms regulating POC accumulation in the northern Adriatic Sea. Three scenario tests (diatom, nanoflagellate and dinoflagellate blooms) using the BN predicted diatom blooms to produce high chlorophyll *a* at the water surface while nanoflagellate blooms were predicted to occur also at lower depths (> 5m) in the water column and to produce lower chlorophyll *a* concentrations. A sensitivity analysis using all available data identified the variables with the greatest influence on POC accumulation being the enzymes, which highlights the importance of microbial community interactions. However, the incorporation of experimental and field data changed the sensitivity of the model nodes ≥25% in the BN and therefore, is an important consideration when combining manipulated data sets in data limited conditions.

Keywords: Bayesian Network; bacteria; phytoplankton; biogeochemical cycling; particulate organic carbon; Adriatic Sea.

1. Introduction

Bayesian Networks (BNs) are being increasingly applied to model complex ecosystem processes through the graphical and probabilistic integration of numerous interacting variables to provide a scientifically informed framework for decision making (Fletcher et al., 2014). The graphical representation of complex interactions between multiple variables can assist in the communication of BNs to end-users thereby facilitating the application of BNs into water resource management practices (McDonald et al., 2015). Although BNs are limited by the inability to model feedbacks that are important in aquatic ecosystem processes unless a computationally demanding dynamic network is developed, they have some benefits that in particular circumstances, such as data limited conditions, can outweigh this limitation (McDonald et al., 2015). A benefit of the BN approach is the ability to iteratively evolve based on the successive incorporation of available and new emerging knowledge of the investigated system into a scientifically informed framework that can be used to investigate probabilistic relationships between variables, make predictions and test scenarios (Lowe et al., 2014; Nojavan et al., 2014). Additionally, the fact that probabilistic dependencies between variables in BNs are explicitly shown supports the communication of the model across disciplines such as management and science, and microbiology and computer science (Fletcher et al., 2014; Levontin et al., 2011). This facilitation of inter-disciplinary collaboration increases the potential for the model to be applied not only within the scientific community but also by a wide ranging end-user community, including environmental managers, regulators and water industries with requiring in-depth understanding of the detailed modelling approach.

Aquatic ecosystems are characterised by complex interactions between variable physical, chemical and biological factors that affect primary production and carbon cycling at different spatial and temporal scales. At the microscale, the structure and strength of bacteria-phytoplankton coupling vary spatially and temporally, and are regulated by nutrient supply (Azam and Malfatti, 2007). The organic matter (OM) pool available in aquatic ecosystems can be conceptualized as a physical continuum of molecules (Verdugo et al., 2004) that spans from colloids and gel particles known as dissolved organic matter (DOC) to particulate organic carbon (POC) aggregates such as marine snow (Alldredge and Cohen, 1987) or even large aggregates of different forms and sizes (mucilage) (Giani et al., 2005 and references

therein). The pathways and rates of dissolved and particulate carbon cycles may be affected by sources, composition and transformations of aggregates in the environment (Turner, 2014 and references therein). The microbial communities and biogeochemical processes of the OM continuum furthermore control the habitat templates and resources for higher trophic organisms (Green and Dagg, 1997). Currently, marine POC formation, accumulation and sedimentation processes are being explored as potential pathways to remove CO_2 from the atmosphere through sequestration *via* photosynthetic fixation of CO_2 into biomass by phytoplankton.

Current models for predicting microbial community changes, such as function based models and bioclimatic models as opposed to a BN approach, have limited ability to link processes to environmental changes in the marine ecosystem and conduct scenario tests on scales relevant for monitoring and management (Larsen et al., 2012). Complex NPHZ-V multi-trophic models (Weitz et al., 2015) have been developed to integrate the complex inter-relationships between viruses, plankton and bacteria but do not reflect the impacts of physio-chemical conditions. Several numerical models have been implemented previously to investigate oceanographic properties linked to atmospheric forces that coincided with large organic aggregates (mucilage) events (Oddo et al., 2005), or to analyse the physical-chemical mechanisms that may regulate aggregation events (Signell et al., 2005) in the Adriatic Sea. Numerical models such as Phytoplankton Aggregation Model (PAM), Snow Aggregate Model (SAM) integrate processes of the microbial cycle but are limited in their application due to their parameterisation requirements and demands on the specialist numerical modeller (Kriest, 2002). The PAM and SAM models aim to characterise the marine snow aggregates by size, density and composition rather than aiming to predict what physical-chemical and biological conditions lead to aggregate events. Cossarini and Solidoro (2008) performed a trophodynamic model to highlight the most important factors for POM accumulation, such as phytoplankton, total phosphorous concentrations, decay rate of particulate organic phosphorous, and mortality rate of bacteria for the Gulf of Trieste. The Mucilage Aggregate Index (MAI) approach was proposed to characterise the aggregate characteristics (size and distribution in the water column) to environmental parameters with correlations (Bragato et al., 2006). These approaches fail to identify and quantify the mechanisms influencing OM aggregates along gradients of physical and chemical attributes that vary spatially and temporally in marine environments. Therefore, there has been a demand for network based

models, such as BNs, that can be applied by scientists and managers to investigate the mechanisms of OM aggregates in data limited conditions (Hurwitz et al., 2014).

The sporadic occurrence and lack of knowledge on the mechanisms of POC accumulation events has resulted in incomplete and limited datasets on the changes within and between ecosystem variables that precede aggregate formation. Integrating multiple data sources, such as expert elicitation with field observations in fuzzy logic approaches, has been commonly used to supplement quantitative information in the development of BNs under data limited conditions (Ban et al., 2014; Isci et al., 2014; Scholton et al., 2012). Combining different sources of *a priori* data, such as combining simulation and field data, can introduce bias and increase uncertainty in the posterior (output) probabilities of BNs that require assessment and in some cases the ranking of data sources (Hamilton et al., 2015). However, the inclusion of manipulative experimental datasets in *a priori* data to fill information gaps in data limited conditions and the consequences on the uncertainty and bias of the resulting posterior probabilities is undetermined.

In this study, a BN was iteratively developed to increase our understanding of the main parameters that effect POC formation in a marine environment using a proof of concept example developed for the shallow and enclosed areas, such as the Gulf of Trieste (GT), northern Adriatic. Several recurring events, either linked to anthropogenic eutrophication or to specific natural conditions, such as hyper-production copious mucus macroaggregates (Giani et al., 2005) have characterised the whole northern Adriatic basin in the recent past. It was shown that the variations in the availability of inorganic nutrients, dissolved organic nitrogen (DON) and dissolved organic phosphorus (DOP) can strongly influence the phytoplankton primary production and the microbial degradation of OM (Cozzi et al., 2004; Danovaro et al., 2005). Under certain poorly understood conditions, the recalcitrant nature of the OM pool combined with slower microbial degradation processes can lead to an increase of the POM pool and formation of large aggregates (Fajon et al., 1999; Malfatti et al., 2014).

Within the model, experimental and field data on microbial activity, including phytoplankton and bacteria communities, was combined with the physical-chemical parameters. Scenario

tests using the set of data available for this case study were conducted to investigate the important processes involved in the POC formation and accumulation. The scenario test assessed the most probable environmental conditions occurring during: (i) a diatom bloom, (ii) a nanoflagellate bloom and (iii) a dinoflagellate bloom. A sensitivity analysis was conducted to assess the causal structure of the BN and the variables that most influence the output probabilities in the three scenario tests. Our hypotheses were that: 1) Phytoplankton community structure and primary production are important factors in POC formation and accumulation; and 2) Bacterial enzymatic activities controlling the transitions between POC and DOC are important factors in POC accumulation. Additionally, we assess the influence of incorporating experimental and field *a priori* data on the posterior probabilities of the BN.

2. Methods

2.1 Study area

The semi-enclosed Gulf of Trieste (GT) is a shallow coastal area (maximal depth of about 25 m) in the northernmost end of the Adriatic Sea. Its oceanographic conditions are affected by water mass exchange with the northern Adriatic at the open boundary, by variable local meteorological conditions that induce a pronounced seasonal cycle of seawater temperature (from 6 ^{o}C in winter to summer peaks of >25 ^{o}C) (Malačič et al., 2006) and by pronounced freshwater inputs of rivers (Cozzi et al., 2012). These physical factors are ultimately reflected in strong seasonal and inter-annual variability in ecosystem structure and functioning, which primarily includes changes in plankton communities and primary production (Fonda Umani et al., 2007; Malej et al., 1995; Tinta et al., 2015). Two seasonal peaks of phytoplankton biomass and abundance regularly occur in the GT: one in spring, being mostly due to the proliferation of nanoflagellates, and the other in late autumn, which is also the highest on the annual scale and is dominated by diatoms (Mozetič et al., 2012). Dinoflagellate abundance represents, with some exceptions, only a small portion of the phytoplankton community (on average around 4%) (France and Mozetič, 2012). At times, phytoplankton dynamics can be altered by exceptional events such as heavy precipitation or enhanced river inputs in summer, resulting in a diatom bloom in July (Malej et al., 1997; Tinta et al., 2015). The bacterial community structure shows the importance of *Alphaproteobacteria* (mainly SAR11), *Gammaproteobacteria* (*Bacteriodetes*, mostly *Flavobacteria*) and *Cyanobacteria* (*Synechococcus*) in GT (Tinta et al., 2015). Less abundant or rare bacterial groups are *Beta-*,

158 *Delta-* and *Epsilonproteobacteria*, *Sphingobacteria*, *Cytophaga*, *Planctomycetes*,
159 *Actinobacteria*, *Verrucomicrobia* and *Deferribacteres.* Seasonal and spatial distribution of
160 bacterial community dynamics is influenced by temperature, freshwater-born nutrients and
161 phytoplankton blooms (Tinta et al., 2015).
162

163 *2.2 Experimental and field data*

164 Two sources of a *priori* data were used to inform the models posterior probabilities
165 (described in detail in the *Model development* section of this paper): a mesocosm experiment
166 and a field study (monitoring), both conducted in the GT.

167

168 An extensive (in terms of biogeochemical parameters analysed) 64-day mesocosm was
169 carried out in October 2007 in order to study carbon and phosphorus fluxes mediated via
170 microbial mechanisms and how interaction between carbon (C) and phosphorus (P) may lead
171 to DOC accumulation and persistence (Malfatti et al., 2014). Natural plankton assemblages
172 (bacteria and phytoplankton while larger herbivores were removed using 50 µm mesh)
173 collected in the south-eastern part of the GT were firstly spiked with nutrients except P at
174 F/10 concentration (Guillard and Ryther, 1962). After, three replicate carboys (P+) received
175 0.5 µM PO_4^{3-} (approx. 10-times higher concentration compared to average phosphate
176 concentration in the sea water) while no PO_4^{3-} was added to the other three (P-) control
177 carboys. The six carboys were incubated *in situ* at 2 m depth. The mesocosm experimental
178 design, parameters sampled and methods used are explained in details in Malfatti et al.
179 (2014). In particular, POC and DOC were measured in samples that were retained on or
180 passed through combusted GF/F filters, respectively, following standard procedures.

181

182 The other set of data originated from a two-year field survey (2009-2010) carried out in
183 fortnightly intervals at the marine field station 00BF (45° 32.93' N, 13° 33.03' E, 1.3 NM off
184 the coast, depth of 22 m), where oceanographic buoy Vida is located, in the south-eastern part
185 of the GT. Samples were collected at the surface (5 m) and near the bottom (20 m) of the
186 water column. The main objective of this study was to examine the seasonal dynamic of the
187 bacterial community of a coastal ecosystem and to investigate potential links between
188 bacterial and phytoplankton community and environmental parameters (for details see Tinta

et al., 2015). Field station 00BF is the same site where the sea water was collected for the mesocosm experiment in 2007 and defines relatively undisturbed open waters of the Gulf. The location also represents one of the longest time-series of the whole GT (Mozetič et al., 2010); some parameters (*e.g.*, dissolved oxygen and chlorophyll *a*) have been continuously measured on a monthly basis from mid 80s onwards. Besides, the location makes part of the grid of sampling stations, which is included in the national monitoring programme and combines, when possible, with stations on the Italian side of GT into a complete coverage of the Gulf's surface.

2.3 Model development

The iterative development of the BN model commenced with a conceptual model to identify the variables (nodes) incorporated into sub-models and conditional relationships between nodes from the available data (Fig. 1). The causal relationships connecting respective nodes were determined by an extensive literature review and expert knowledge of the authors of this paper on the considered processes in the northern Adriatic Sea. A BN was developed from the conceptual model in Fig. 1 using the modelling software Netica 4.16 (Norsys Software Corporation, 2010).

The conceptual model was initially developed into a network model in which the conditional probabilities were derived from cause and effect relationships (Fig. 1). This model was assessed with sensitivity analysis to investigate the propagation of probabilities through the network structure. The bacterial and phytoplankton sub-models were then further developed to include the taxonomy of phytoplankton and bacteria community structure using naïve Bayesian network (NBN) relationships. The NBN structure assumes independence between each taxonomical variable in the network (Flores et al., 2014). This NBN approach is considered a more simplistic, hence naïve, representation of environmental relationships than the cause and effect approach in defining the conditional structure of the model (Costa et al., 2013). Despite this simplified assumption NBN approaches have strong mathematical foundations and are effective in large, complex models with data limited conditions or for unstructured data (Li and Li, 2013; Xu and Ma, 2014). The BN with both causal and naïve structure was then assessed again with a sensitivity analysis and the results are outlined in

this paper. The development of the bacteria and phytoplankton community composition sub-models into the NBN structure provides important information on the abundance of each taxonomical class rather than only the dominant taxa in the initial cause and effect model structure that was developed (Fig. 1). A sensitivity analysis was used to assess model propagation through the final network structure and identify any insensitive or poorly informed nodes which could indicate to problems in the network structure.

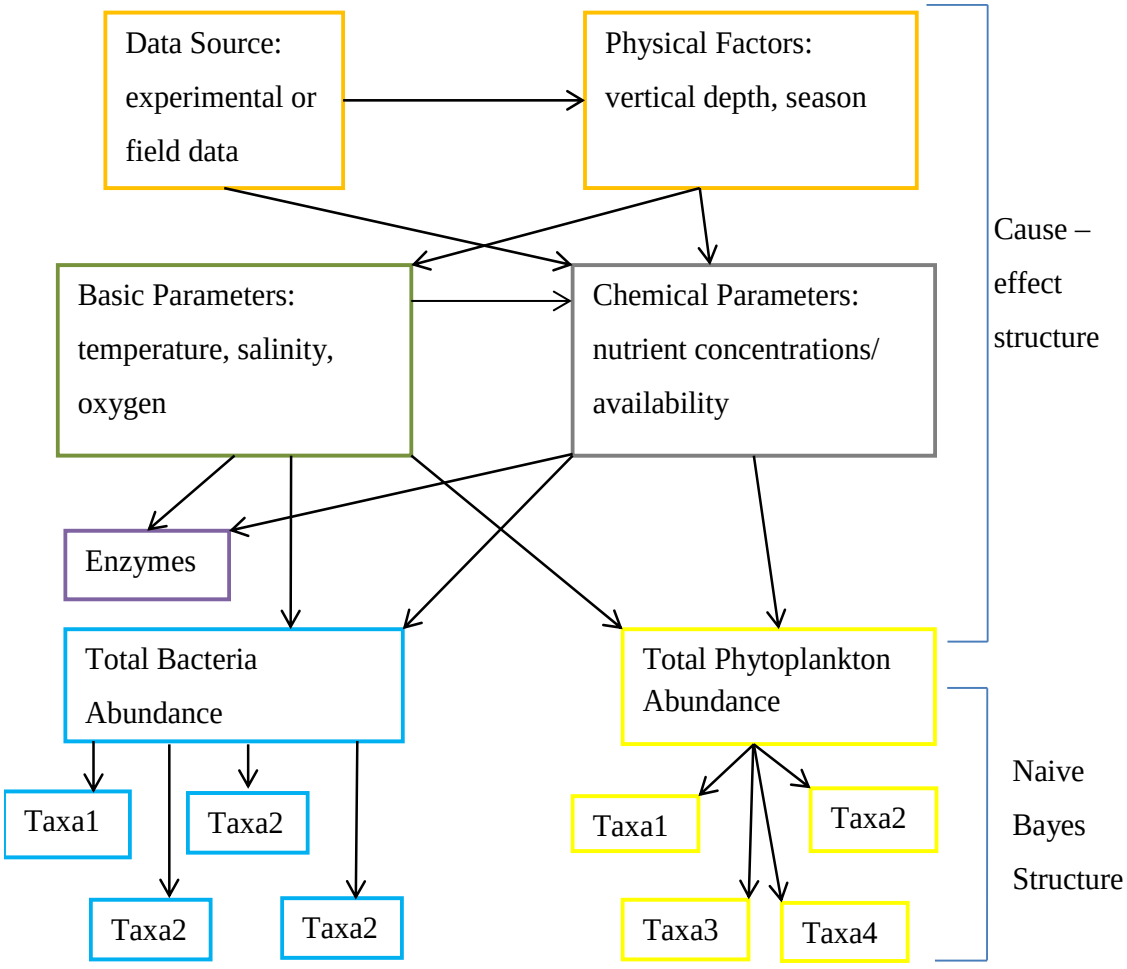


Fig. 1 Conceptual model of the main causal relationships between sub-models of nodes in the network describing processes and process interactions leading to the higher POC concentrations. The arrows indicate conditional dependencies between sub-models.

In our study, the physical-chemical and biological parameters that might be relevant for formation and accumulation of POC were integrated into a network model. The *a priori* data

used to calculate posterior probabilities for each node in the network combined data from two-year field survey (2009-2010) with a mesocosm experiment (described in detail in the *Experimental and field data* section of this paper) into one case file (Supporting Information 1). The BN is developed with a data node that can be used to distinguish between the data sources (Fig. 1). The inclusion of experimental data to supplement the field data was important to fill information gaps that exist in the 2-year monitoring campaign, such as data on enzyme availability and nutrient thresholds. Therefore, the experimental data are important to inform the relationships among variables in the model and quantify trends that the biweekly to monthly monitoring scheme may not detect. The posterior probabilities derived from only field data and only experimental data are assessed using the data node embedded in the network and a sensitivity analysis conducted to investigate changes in influences between variables between the two data sources.

The BN in our study was developed to the best-practice principles outlined in McDonald et al. (2015) as a proof of concept for modelling microbial community interactions in aquatic environments using BNs. Therefore, it is intended that the model is in the initial phase of construction to assess the approach and will be developed in the future prior to being applied to ecosystem management. The states for each node (Mild, Mean, Moderate, Maximum) were defined by percentiles of all available data which is a common method in data limited conditions (Pollino et al., 2007). The ecological importance of each node included in the network is outlined in Supporting Information 1. Conditional probability tables (CPTs) were calculated from the data using the Expectation Maximization (EM) algorithm due to the limited number of cases (data points) for each node (cases are provided in Supporting Information 1; CPTs and further model configuration is available from the lead author on request). The CPTs produced in this unconditioned model represent the base case (the parameterized model prior to a user defined scenario being entered) or the probabilities based on all possible outcomes of the *a priori* data (McDonald et al., 2016).

2.4 Scenario testing

The complex interactions between mechanisms regulating POC accumulation and microbial community structure in the unconditioned BN model developed have been furthermore tested

in three user-defined scenarios. Assuming that a chosen model structure is accurate, scenario tests (where the model is conditioned to have a specific CPT outcome for at least one node) can provide information on ecosystem responses under specific conditions (Mantyka-Pringle et al., 2014; Van Grieken et al., 2013). The first scenario test investigated the ecosystem responses under a high abundance of diatoms, the second scenario test a high abundance of nanoflagellates, and the third scenario test included a high abundance of dinoflagellates. Scenario tests were conducted by setting the abundance node for the phytoplankton group in question (nanoflagellates, dinoflagellates or diatoms) to 100% probability of the state representing the highest possible concentrations occurring and the dominant phytoplankton node finding to 100% probability of occurrence for the same group being investigated in the scenario test (*i.e.* diatom, dinoflagellate or nanoflagellate, respectively) (Supporting Information 2). Thereby, predicting the probable influence a bloom event of a particular phytoplankton group being investigated on the microbial community and POC. The predictions for investigating the ecosystem responses in this study were conducted using both the forward and backward propagation techniques (McDonald et al., 2015). These scenario tests were used by the authors to investigate both the interactions between phytoplankton community composition and the physical-chemical factors regulating carbon accumulation and degradation processes in the GT.

2.5 Microbial mechanisms

Microbial interactions with POC aggregate formation were investigated using the scenario testing technique outlined above and compared to the posterior probabilities in the unconditioned model. The microbial mechanisms were investigated by conducting a scenario test in which the POC node was set maximum state threshold and all parent nodes remained unconditioned during the scenario test (Supporting Information 2). Thereby, the scenario test investigated the most probable environmental and microbial conditions present under the highest POC conditions in the *a priori* data were assessed using the BN model.

2.6 Sensitivity analysis

The sensitivity analysis, combined with the scenario testing, identified the nodes that are most sensitive to changes in the posterior probabilities (calculated in the conditional

probability tables (CTPs)) with different outcomes in network. A sensitivity analysis was conducted on the nodes of interest such as POC, phytoplankton and bacteria abundance. All parent nodes remained unconditioned during the sensitivity analysis. The variance reduction (VR) method in the Netica software was used to calculate the sensitivity between nodes in the network.

Sensitivity analysis was furthermore used to investigate the mechanisms regulating POC increase based on the model being informed by: (i) all the *a priori* data available, (ii) only the experimental data and (iii) only the field data available. The changes in mechanisms of POC accumulation identified by the sensitivity analysis were investigated by conducting scenario tests on the unconditioned model for the all available data case study. The different data sources were then investigated by setting data node to 100% probability of field data or 100% probability of experimental data being used to derive the model outputs. Inferring differences between the sources of the *a priori* data (field and experimental) aimed to identify the variability that may be introduced into the model through the incorporation of manipulated experimental data to fill information gaps in field (monitoring) data.

3. Results

The output probabilities in the unconditioned BN predict a mean POC concentration to be the most likely (48.0%) under all possible outcomes (base case) from the *a priori* data (Fig. 2; Table 1). The most probable outcome for chlorophyll *a* concentrations was predicted to fall within maximum threshold (26%) or within mean threshold (21.4%). The abundance of total phytoplankton (25.7%) and rates of primary production (41.7%) were predicted to be in the mean state range. The concentrations of DOC are predicted to fall within the minimum threshold range (51.8%). The dominant bacteria order is predicted to be SAR 11 (38.9%) based on all possible outcomes from the *a priori* data (Fig. 2). Nanoflagellates have the highest probability of being the dominant phytoplankton group (40.5%).

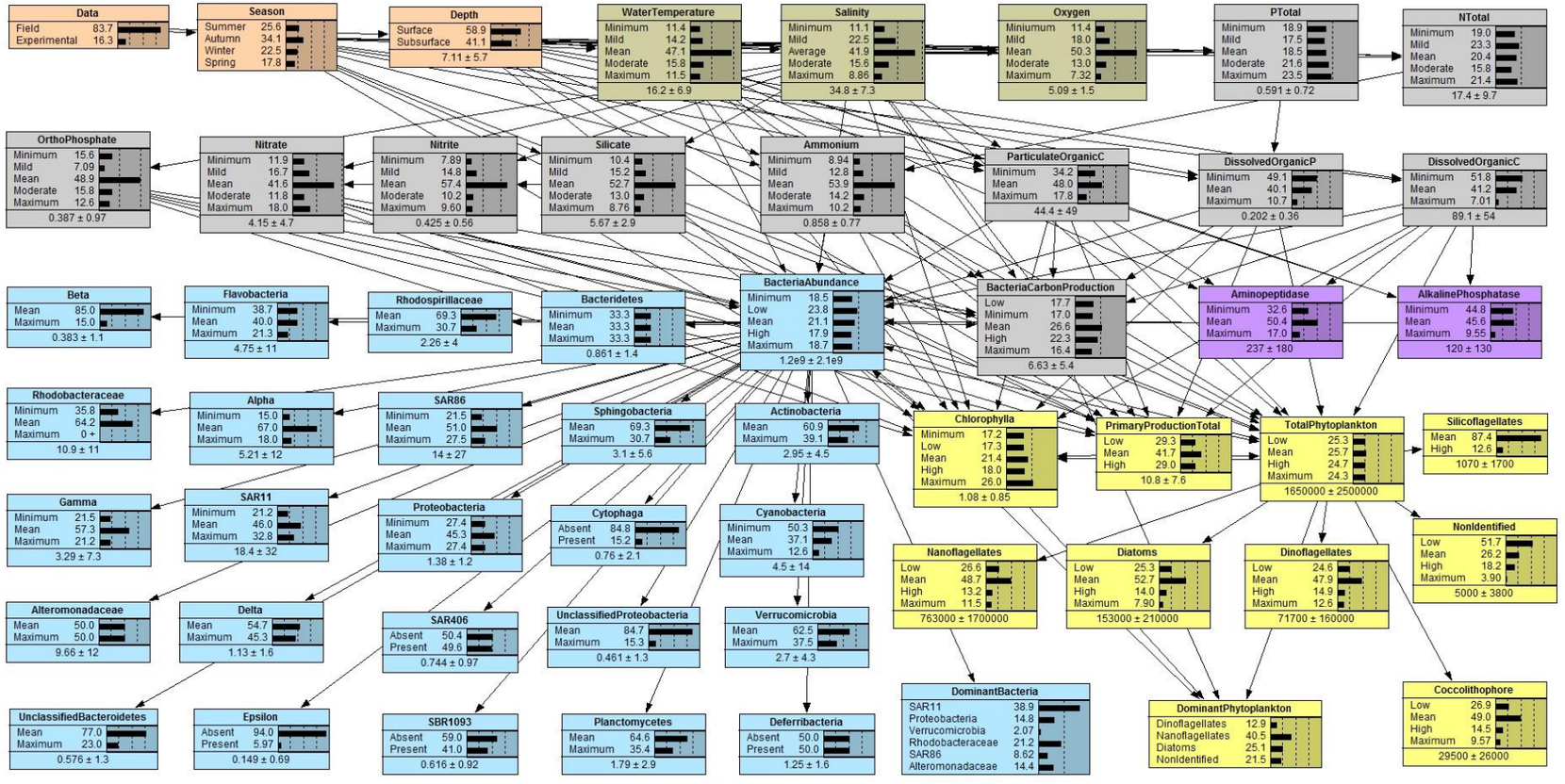


Fig. 2 The unconditioned BN developed for investigating microbial mechanisms that lead to POC accumulation in marine environments. The network comprises of the spatial and temporal variables (orange), the basic physical-chemical parameters (green), the inorganic and organic nutrients (grey), enzyme activities (purple), bacteria community composition (blue) and phytoplankton community composition (yellow). Bottom layer (20 m) is regarded here as the subsurface layer.

330 **Table 1.** The predicted probabilities of states for key physical and chemical drivers of POC accumulation in each of the scenarios (to 1 decimal
331 place).

Node	State	All probable outcomes	High Diatom abundance	High Dinoflagellate abundance	High Nanoflagellate abundance	Maximum POC
Season	Summer	25.6	27.5	24.5	26.1	9.7
	Autumn	34.1	36.4	33.9	36.9	35.0
	Winter	22.5	17.8	23.9	16.1	17.1
	Spring	17.8	18.2	17.7	20.9	38.1
Temperature	Minimum	11.4	10.6	11.1	11.4	7.8
	Mild	14.2	12.6	14.4	11.1	27.4
	Mean	47.1	46.5	46.9	50.5	46.2
	Moderate	15.8	17.1	15.7	15.8	13.0
	Maximum	11.5	13.2	11.9	11.2	5.5
Depth	Surface	58.9	59.1	61.8	55.8	71.3
	Bottom	41.1	40.9	38.2	44.2	28.1
Dissolved oxygen (DO)	Minimum	11.4	12.4	11.3	12.1	5.69
	Mild	18.0	19.0	17.7	19.5	22.1
	Mean	50.3	48.2	51.2	47.1	53.6
	Moderate	13.0	13.0	13.0	13.1	15.3
	Maximum	7.3	7.4	6.8	8.1	3.4
Ammonium	Minimum	8.9	8.62	10.5	7.3	8.96
	Mild	12.8	12.9	13.8	11.2	13.5

	Mean	53.9	53.9	53.0	53.7	62.8
	Moderate	14.2	14.3	13.6	16.4	5.24
	Maximum	10.2	10.3	9.08	11.4	9.53
Ortho-phosphate	Minimum	15.6	16.6	16.0	15.1	10.9
	Mild	7.09	7.4	7.6	6.4	11.2
	Mean	48.9	46.4	47.0	51.8	47.5
	Moderate	15.8	15.4	15.9	15.3	17.1
	Maximum	12.6	14.2	13.5	11.3	13.4
Total Phosphorus	Minimum	18.9	19.1	18.8	19.3	15.3
	Mild	17.5	17.4	17.4	17.6	15.0
	Mean	18.5	18.2	18.4	18.6	15.9
	Moderate	21.6	21.5	21.6	21.3	14.9
	Maximum	23.5	23.8	23.8	23.2	38.8

3.1 Scenario testing

The BN model predicts increased probability of POC accumulation in autumn in the unconditioned network from the *a priori* data (34.1%), that increased in the nanoflagellate bloom scenario test (36.9%) and in the diatom bloom scenario test (36.4%) but decreased in the dinoflagellate bloom scenario test (33.9%), (Table 1; Supporting Information 2). Diatom abundance is predicted to increase with water temperatures in the high 17.1% or maximum 13.2% node states. The model output predicts the vertical distribution of diatom (59.1%) and dinoflagellate (61.8%) with higher abundance in the upper water column, while nanoflagellates are more evenly distributed between the surface (55.8%) and bottom layer (44.2%).

3.2 Microbial mechanisms

The maximum POC scenario is predicted to occur in spring (38.1%) at the surface (71.3%) (Table 1; Supporting Information 2). The maximum concentrations of total phosphorous (38.8%) in the *a priori* data are predicted to occur during POC accumulation events. During the maximum state POC concentrations the probability of low DO concentrations (in the mild (22.1%), mean (53.6%) and moderate (15.3%) node state ranges) increases from the probabilities at base case in the unconditioned model.

Elevated dinoflagellate abundance is reflected in increase of the chlorophyll *a* concentration within the maximum threshold state of 35.5%, an increase from 26% based on all probable outcomes (base case) in the unconditioned network, 11.2% in high nanoflagellate conditions and 30% in high diatom conditions (Table 2). Predictions for phytoplankton abundance falling within the maximum threshold state was greatest in the diatom scenario (50.8%), which increased from 26% in the unconditioned network. The probability of a maximum phytoplankton abundance decreased with a nanoflagellate (13.7%) bloom and a dinoflagellate (5.7%) bloom. The probability of primary production in the high range increased from 29% in the unconditioned network to 30.1% with high diatom abundance and 30.7% with high nanoflagellate abundance. The probability of bacteria abundance occurring within the maximum node state range also increased from 18.7% in the unconditioned network to 23.1% in the diatom scenario test and 19.8% in the nanoflagellate scenario test. The maximum

365 concentrations of POC were expected to increase from 17.8% to 18.7% in the diatom
366 scenario and 18.4% in the dinoflagellate scenario for POC (Table 2; Supporting Information
367 2). The probability of higher DOC concentration increased from 7.07% to 7.8%, based on all
368 probable outcomes in the unconditioned network, with high diatom abundance and even more
369 (7.2%) with high dinoflagellate abundance.

370

371 **Table 2.** The BN output probabilities (%) for key variables that indicate changes in biotic community structure and carbon accumulation (to 1
372 decimal place). The states with the highest probable outcome are in bold (Full model outputs are provided in Supporting Information 2).

Node	**State**	**All probable outcomes**	**High diatom abundance**	**High dinoflagellate abundance**	**High nanoflagellate abundance**	**Maximum POC**
Chlorophyll *a*	Minimum	17.2	15.1	19.8	16.8	17.9
	Low	17.3	11.2	24.5	14.8	18.9
	Mean	21.4	6.4	0	**42.9**	20.3
	High	18.0	**37.3**	20.3	14.3	17.7
	Maximum	**26.0**	30.0	**35.3**	11.2	**25.2**
Phytoplankton abundance	Low	25.3	0	0	0	**25.5**
	Mean	**25.7**	18.4	4.0	8.1	24.8
	High	24.7	30.8	**90.3**	**77.4**	24.5
	Maximum	24.3	**50.8**	5.7	13.7	25.2
Net Primary production	Low	29.3	30.5	28.5	30.4	30.8
	Mean	**41.7**	**39.5**	**43.3**	**38.9**	**38.6**
	High	29.0	30.1	28.1	30.7	30.6
Bacterial abundance	Minimum	18.5	15.1	21.8	15.7	19.7
	Low	**23.8**	17.1	**27.5**	21.1	**21.9**
	Mean	21.1	**27.6**	18.2	20.3	20.3
	High	17.9	17.0	15.5	**23.1**	18.4
	Maximum	18.7	23.1	17.1	19.8	19.7
POC	Minimum	34.2	34.8	33.2	36.0	0

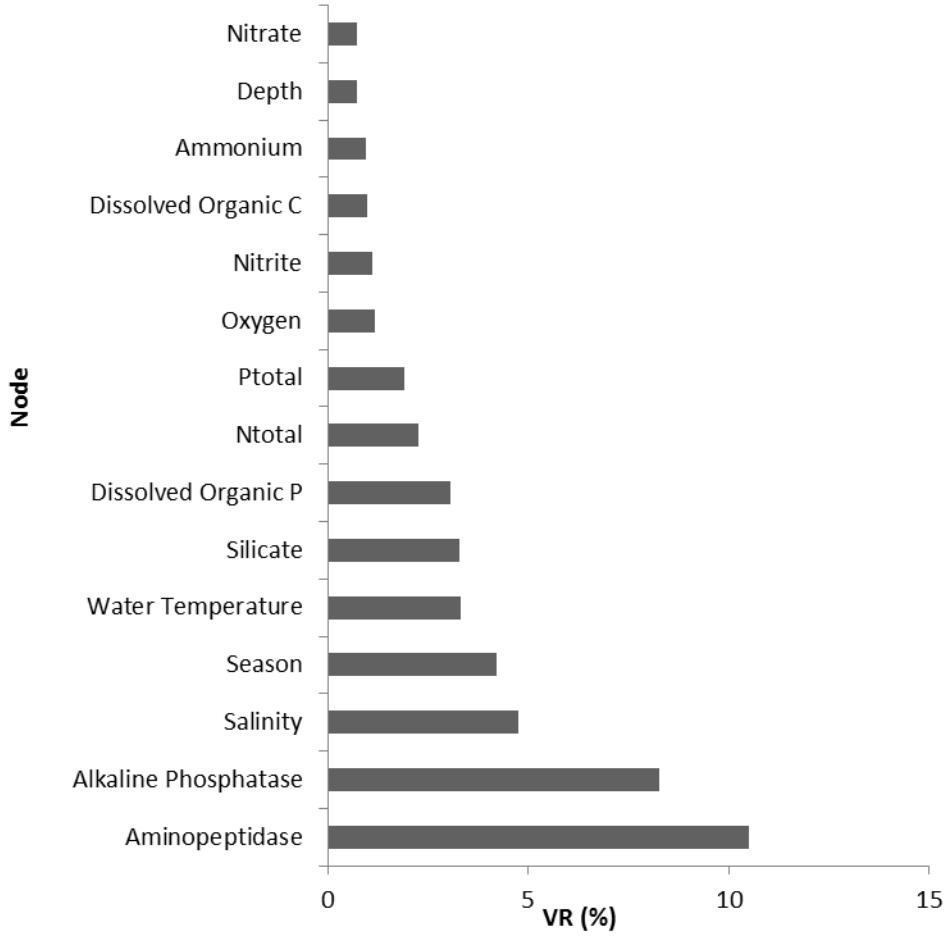
	Mean	**48.0**	**46.6**	**48.4**	**46.6**	0
	Maximum	17.8	18.7	18.4	17.4	100
DOC	Minimum	**51.8**	**55.2**	**50.4**	**55.7**	**44.9**
	Mean	41.2	37.0	38.9	37.4	27.3
	Maximum	7.0	7.8	7.2	6.9	27.7

373

The BN predicted chlorophyll *a* concentrations in the maximum state range and low total phytoplankton abundance (25.5%) to be the most probable for POC concentrations in the maximum node state (Table 2; Supporting Information 3). The most likely bacteria abundance during maximum POC concentrations is predicted to be within the low state concentrations (21.9%). The minimum node state concentrations of DOC were the most probable (44.9%) to occur during events where the POC is within the maximum concentrations.

3.3 Sensitivity analysis

The variables with the greatest influence on the probability of POC accumulation in the model were aminopeptidase (10.5%) and alkaline phosphatases (8.3%) (Fig. 3; Supporting Information 3). The POC output probabilities were sensitive to salinity (4.7%). The seawater temperature (3.3%) and silicate (3.3%) were also key factors influencing the probability of POC accumulation in the system that may account for some of probabilistic changes in the community composition scenarios investigated in this paper. The dominant phytoplankton class probabilities were most sensitive to changes in the chlorophyll *a* (8.7%) and phytoplankton abundance (3.3%) nodes. The probabilities of phytoplankton community composition nodes, such as coccolithophorids (1.5%) and dinoflagellates (1.4%), were identified as key variables influencing the model output probabilities for dominant phytoplankton class. Probabilities for the bacteria abundance node were most sensitive to changes in the probabilities of the dominant bacteria (53.2%) node (Supporting Information 3). Bacteria community composition nodes, such as *Flavobacteria* (38.6%) and SAR11 (37.8%), were also key nodes influencing the probabilities of the bacteria abundance.



398

399 **Fig. 3** Sensitivity analysis indicating the variables that have the greatest influence on the POC
400 node based on all available data being used to inform the model. Nodes are provided for
401 values up to ≥1% VR change.

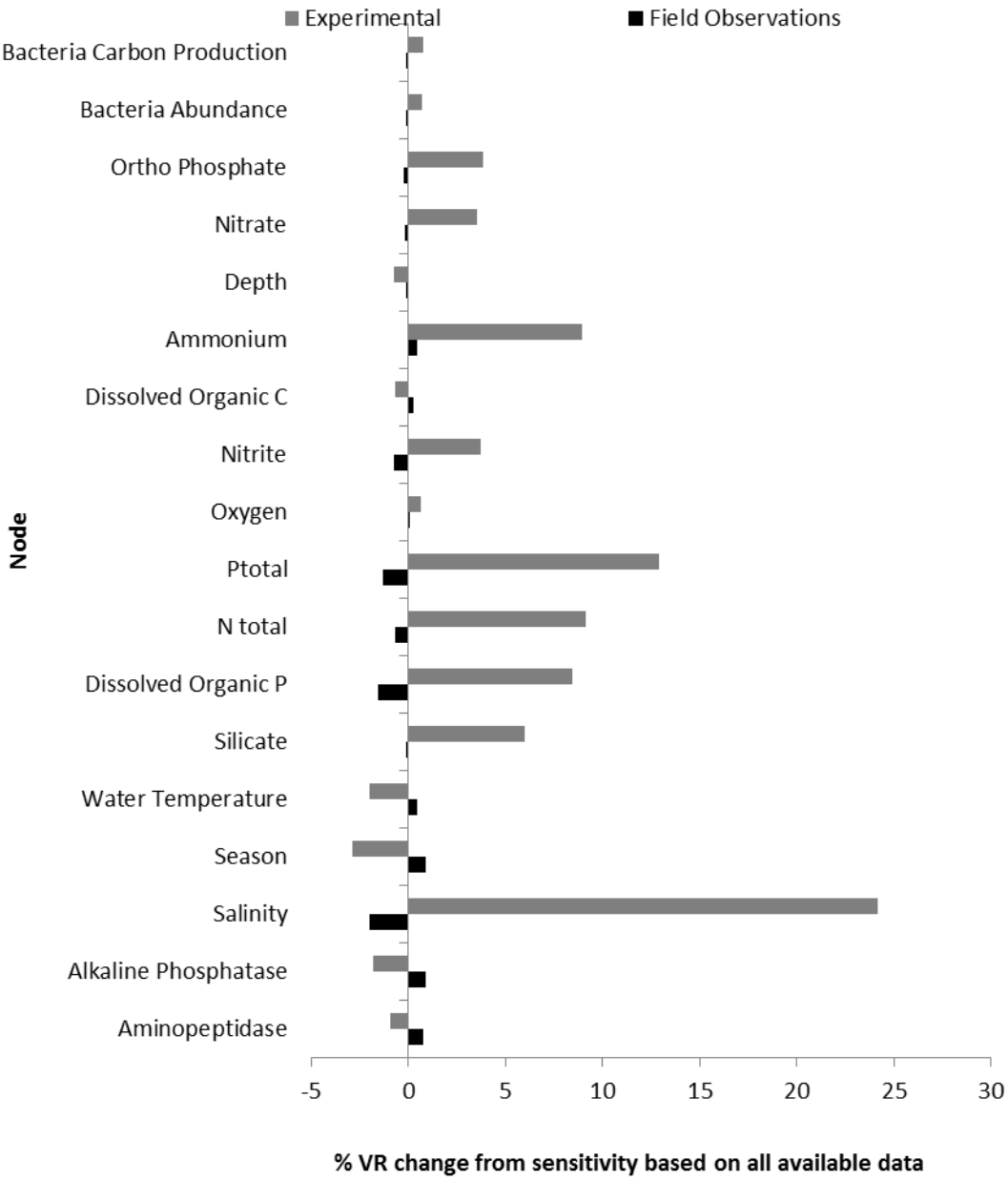
402

403 *3.4 Microbial mechanisms by data source*

404 The embedded data node in the BN structure distinguished between posterior probabilities
405 derived from manipulated experimental and field data. The posterior probabilities of nodes
406 populated with experimental data predict higher temperatures (Maximum 18%), lower
407 salinities (Mild 54%), lower oxygen (Mild 29%), and increased DOC (Maximum 24%) and
408 DOP (Maximum 28%), than if the model was populated by only field data (Supporting
409 Information 4). Additionally, the posterior probability of the season node predicts the seasons
410 that the experiments were conducted in (autumn 81% and winter 19%) when informed by
411 only the experimental data. The BN populated with experimental data predicts higher bacteria
412 abundance (21%), increased total phytoplankton (25%) and a shift in the dominant bacteria

(*Rhodobacteraceae* 20%). The posterior probabilities of nodes informed with field data predict a broader range of probabilities among all node states than the probabilities from only experimental data. For example, lower bacteria abundance (Maximum 19%), and increased ammonium (11%), silicate (10%), nitrite (10%), nitrate (15%) and orthophosphate (13%) was predicted in the BN informed by only the field data.

The nodes that had the greatest influence over the outcomes of POC in the model varied between whether all data, only experimental data or only field data were used to inform the posterior probabilities. The variables with the greatest influence on POC using all available data to inform the probabilities were: aminopeptidase (10.5%) and alkaline phosphatases activities (8.3%), salinity (4.7%), and season (4.2%). The variables with the greatest influence on POC using only field data were: salinity (28.9%), total phosphorus (14.8%), DOP (11.5%), and total nitrogen (11.4%). The variables with the greatest influence on POC using only experimental data were: aminopeptidase (11.3%) and alkaline phosphatases activities (9.2%), season (5.1%), and water temperature (3.8%). Changes in the sensitivity of the POC node from the base case probabilities (all available data in the unconditioned model) were greatest for the season node which increased by 0.9% and the salinity node which decreased by 2% in the field data (Fig. 4). In the experimental data the sensitivity of the POC node increased 24.16% to salinity and season decreased by 2.9% from the base case with all available data.



433

434 **Fig. 4** The % deviation from the VR (%) of the model at base case (all probable outcomes)
435 when the model is informed by either only field data, or only experimental data. Nodes are
436 provided for values up to ≥1% VR change.

4. Discussion

The integration of numerous sources of data is a key strength of the BN approach for data limited conditions (Ban et al., 2015; Li et al., 2010) when the node states, network structure and learning algorithms are carefully selected and applied (Lucena-Moya et al., 2015; Lui et al., 2007). Data from experimental ecosystems provide valuable information for predictive models on ecological thresholds that have not been exceeded and thus, are not detected in field datasets (Perlinski et al., 2014; Van Dam et al., 2014). However, including manipulated experimental data into predictive models can skew the output probabilities away from trends observed in the nature. By embedding a data source node into the network the posterior probabilities and interactions between variables can be investigated by end-users without manipulative data if bias is suspected. End-users can then determine the confidence in the model outputs and make informed decisions accordingly. Therefore, by incorporating manipulated date in the BN the interactions between variables in the predictive model can be informed from all available knowledge of the system in data limited conditions.

Understanding the possible uncertainty and bias in the *a priori* data, such as manipulated experimental data, is essential for managers and scientists to make informed interpretations of the model predictions. Salinity had the greatest variability in the sensitivity of the POC node and was more important in the experimental data (25% VR difference from the unconditioned model probabilities) and less important in the field data (-3% VR difference from the unconditioned model probabilities). This variability in the sensitivity of the POC node, and therefore POC aggregates, to salinity could be influenced by the lack of seasonal freshwater fluxes and depth profile in the simplified representations of the environment in the mesocosm experiments (Puddu et al., 1997; Monticelli et al., 2014). The sensitivity of the POC node to nutrient availability and turnover was also higher in the experimental data to the field data and highlights the influence of the nutrient enrichment on POC and system function. The enzymatic activity (alkaline phosphatase and aminopeptidase) remained among the most influential nodes on POC in both the manipulated and experimental data. However, alkaline phosphatase and aminopeptidase dropped from being the most influential nodes on POC when the model is informed by the manipulated experimental data which could be a result of the lack of data informing the nodes under the scenario. Similarly to the inclusion of

simulation or qualitative data into BNs, the inclusion of experimental data in the *a priori* data is an acceptable method to fill information gaps or provide information on event that are yet to occur (McDonald et al., 2015). In our BN, the manipulated experimental data provided essential information on P thresholds for POC aggregates and filled information gaps on interactions between nutrient availability, microbial community structure and enzymatic activity. However, assessing the differences in the sensitivity of a target variable to the different data sources is important to effectively interpret the model predictions, particularly in relation to regulatory mechanisms.

Overall, the posterior probabilities of our BN under the scenarios presented in this paper support the current understanding of coastal ecosystem functioning (e.g., Malej et al., 1995; Mozetič et al., 2012; Malfatti et al., 2014; Tinta et al., 2015) such as node states in the high and mean chlorophyll *a* concentrations will most probably develop during a diatom (37%) and nanoflagellate (43%) bloom, respectively. In the GT, diatoms have been observed to be responsible for the highest seasonal blooms, which usually occur in autumn (October-November) and recently also in mid-summer (June-July) (Mozetič et al., 2012; Tinta et al., 2015). These observations clearly support the BN prediction of a high probability of high diatom abundance in autumn in the surface layer and indicate that the model is propagating probabilities from the *a priori* data well in the current absence of validation. In general, diatoms are known to thrive under conditions of elevated nutrients (Fawcett and Ward, 2011) and grow at sufficiently high rates to maintain a major contribution to the biomass (Goericke, 2002). In a recent study Talaber et al. (2014) found abundance of diatoms in the surface layer most closely related to high concentrations of total inorganic nitrogen and slightly less to silicate, which corresponded to periods of diatom abundance peaks in late autumn and spring – summer. Tamše et al. (2014) suggested that, besides mixing of waters of different origin, phytoplankton uptake controlled the distribution and isotopic composition of nitrate in the marine system and was more extensive in spring, while in autumn ammonium, not nitrate, was the dominant source for phytoplankton. Indeed, the posteriror probabilities of the POC node in our BN was also more sensitive to changes in ammonium than nitrate.

Nanoflagellates, which are on the annual basis prevailing and most abundant group in the GT, have not been identified to contribute as much as diatoms or dinoflagellates to chlorophyll *a*

biomass. The model prediction that mean chlorophyll *a* concentrations will most probably occur during periods of high nanoflagellate abundance, therefore reflects the real situation of the GT (Mozetič et al., 2012) and of other temperate coastal areas, e.g. Gulf of Naples (Ribera d'Alcalá et al., 2004) and western Black Sea (Yunev et al., 2007). Lastly, the prediction of the model that the maximum state of chlorophyll *a* is most probably achieved during the third scenario of high dinoflagellate abundance is overestimated due to an unusally high abundance of dinoflagellates in the year, which was used to populate the BN.

Our model also predicted the highest probability (44.2%) that a nanoflagellate bloom will develop in the deeper water column layer (>5m) than a diatom bloom (<5m). An explanation could be the fact that flagellates can perform active swimming, which permits these organisms to access the water layer with an adequate quantity of inorganic nutrients, thereby improving their retrieval (Smayda, 1997). It, however, failed to detect dinoflagellates using the same advantageous characteristic, vertical mobility, according to the probability (38.2%) that a dinoflagellate bloom will develop in the bottom layer (>5m).

In the BN model of this study, probabilities for the bacteria abundance node were most sensitive to changes in the probabilities of the dominant bacteria (53%) node, with *Flavobacteria* (37%) and SAR11 (38%) as key nodes with high VR value. These observations clearly support our previous results and significant relationships between diatom blooms and shift in bacterial community composition within *Alphaproteobacteria* (from SAR11 to *Rhodobacteraceae*) and increase of *Gammaproteobacteria* (within which mostly *Alteromonadaceae*, SAR86 and *Vibrionaceae*) (Tinta et al., 2015) in accordance with others (Gilbert et al., 2012; Teeling et al., 2012). *Gammaproteobacteria* appear to be dominant colonizers of diatom detritus (Bidle and Azam, 2001) and marine snow aggregates (DeLong et al., 1993). High variability in bacterial and phytoplankton community composition has been observed in the aggregates during periods of mucilage formation in the northern Adriatic (Najdek et al., 2002).

Together with complex network structure, an importance of enzyme activity during large aggregates events in the GT has been observed (Del Negro et al., 2005; Ivančić et al., 2009;

Turk et al., 2010), changing quality of the DOC (Faganeli et al., 1995; Giani et al., 2005; Malfatti et al., 2014). These observations support our BN model output, since the POC node is most sensitive to changes in the bacterial aminopeptidase and alkaline phosphatase activities nodes in model at base case (all probable outcomes) and the probabilities predicted from only field and only the experimental data. Bacterial extracellular enzymes such as aminopeptidase, lipase, glucosidase, N-acetylglucosaminidase in addition to alkaline phosphatase are important catalysts in the degradation of POC to DOC (Smith et al., 1995). The BN outputs highlight that further investigation of interactions between enzyme availability and bacteria community composition and abundance should be conducted to quantify the relationship as a mechanism in the degradation of aggregates.

The explicit quantification of uncertainty in the model output probabilities and parameterization of the node states from the *a priori* data will be quantitatively assessed as the model is updated and further developed. Qualitative indications of the uncertainty in our model were investigated through the iterative model development (set of alternate models) and the validation posterior probabilities and node sensitivity against known behaviour of the system (Melbourne-Thomas et al., 2012). Uncertainty is expected to arise from the inclusion of field and manipulated experimental data and the lack of feedback loops, such as the microbial loop, in the model structure. Additionally, the short time series and data limited conditions could inhibit the model from detecting long term trends or highly sporadic events outside the available data. Therefore, until our model is updated and uncertainty within our model is numerically quantified, the posterior probabilities have limited ability to inform decision making processes. However, our model outputs can indicate the key nodes, variable interactions and information gaps that are important to direct the future development of the model and scientific investigations of POC aggregate events.

The BN framework developed in this study demonstrated exceptional potential to be developed into a model that can be applied to investigate POC accumulation and microbial dynamics in marine environments. The posterior probabilities captured the statistical trends in the *a priori* data, reported in Tinta et al. (2015) and Malfatti et al. (2014), through the propagation of probabilities in the network indicating the network structure adequately represents current knowledge on ecosystem interactions. Furthermore, the model structure

was sensitive to changes in the CPTs and contained a very small number insensitive or poorly informed nodes particularly for the model size and data availability. Therefore, further development of the model may provide valuable information for managers and scientists on the microbial interactions that regulate POC accumulation and are lacking in current models used for characterising aggregate formation. However, the model was developed as a proof of concept that inherently included two notable limitations. Data availability limitations currently existing in the model can in future be overcome by updating with additional data and conducting an uncertainty analysis as it becomes available. The inclusion of incomplete datasets collected over sporadic timeframes is another key benefit of BNs in predicting data limited environmental events (Ban et al., 2014; Metcalf et al., 2014). A further limitation is that all field data used to inform the BN model was obtained from one location in the GT. Data from additional locations, particularly from the Italian coast near the largest river inflows, may be added in the future for the model to derive probabilities across the spatial extent of the GT. Despite these limitations the BN created in this study advances previously developed complex numerical models such as trophodynamic models (Cossarini and Solidaro, 2008) by integrating physical, chemical and biological variables to investigate the mechanisms for marine POC accumulation in a framework that has the potential to be used by managers, scientists and stakeholders.

A benefit of the BN approach is the adaptability of the approach to be integrated into adaptive management strategies and frameworks (McDonald et al., 2015). The BN presented in this study could allow scientists and managers to identify and prioritise research on information gaps on the poorly understood and complex relationships between the chemical parameters and microbial activity. Our BN model identifies that carbon and nitrogen availability (and turnover) is an important indicator of POC aggregate events, and interactions between enzyme activity and bacterial community composition is important in regulating POC conditions in marine ecosystems. However, little data is available on enzyme activity and including enzymes in microbial monitoring schemes is currently not a common practice. Consequently, quantifying the dynamic relationships between bacteria community structure and enzyme activity remains poorly understood in the lead up to, during and decomposition of POC aggregates. Our BN structure is a transparent and scientifically informed framework to identify and targeted variables for managing POC aggregates that has the potential to be implemented across international borders such as the Adriatic Sea. Further updating with

information from additional sites the model framework developed in this study could be coupled with a GIS interface that scientists and managers could integrate into informed decision making frameworks (Kocabas et al., 2012; Stelzenmuller et al., 2013). Therefore, a BN approach that integrates physical, chemical and biological factors into a decision making framework is an important step forward in predicting and managing POC aggregate events in the future.

Acknowledgements

Financial support for this research was provided by the European Union's Seventh Framework programme for research, technological development and demonstration under grant agreement no. 607150: Ecohydrological Interfaces as Critical Hotspots for Transformations of Ecosystem Exchange Fluxes (INTERFACES). The authors are grateful to the European Union's Marie Curie Actions for funding this research.

References

Alldredge AL, Cohen Y. (1987). Can microscale chemical patches persist in the sea? Microelectrode study of marine snow, fecal pellets. *Science,* 235(4789):689-91.

Azam F, Fenchel T, Field JG, Gray, JS, Meyer-Reil, LA, Thingstad, F. (1983). The ecological role of water-column microbes in the sea. *Marine Ecology Progres Series,* 10:257-263.

Azam F, Malfatti F. (2007). Microbial structuring of marine ecosystems. *Nature Reviews Microbiology,* 5:782-791

Ban SS, Pressey RL, Graham NAJ. (2014). Assessing interactions of multiple stressors when data are limited: A Bayesian belief network applied to coral reefs. *Global Environmental Change*, **27**: 64-72. doi: 10.1016/j.gloenvcha.2014.04.018.

Bidle KD, and Azam F. (2001). Bacterial control of silicon regeneration from diatom detritus: significance of bacterial ectohydrolases and species identity. *Limnology and Oceanography* 46: 1606–1623

Bragato D, Bergamasco A, Bianchi E. (2006). Quantifying macro-aggregate distributions using a new index (MAI). *Journal of Marine Systems*, **61**(1-2): 1-12. doi: 10.1016/j.jmarsys.2006.01.001.

Brigolin D, Lovato T, Rubino A, Pastres R. (2011). Coupling early-diagenesis and pelagic biogeochemical models for estimating the seasonal viability of N ad P fluxes at the sediment-water interface: Application to the northwestern Adriatic coastal zone. *Journal of Marine Systems*, 87(3-4): 239-255. doi: 10.1016/j.jmarsys.2001.04.006.

Brock, T, Hammers-Wirtz, M, Hommen, U, Preuss, T, Ratte, H-T, Roessink, I, Strauss, T., Brink, P. (2015). The minimum detectable difference (MDD) and the interpretation of treatment-related effects of pesticides in experimental ecosystems. *Environmental Science and Pollution Research*, **22**(2): 1160-1174. doi: 10.1007/s11356-014-3398-2.

Celussi M, Del Negro P. *(2012).* Microbial degradation at a shallow coastal site: Long-term spectra and rates of exoenzymatic activities in the NE Adriatic Sea. *Estuarine, Coastal and Shelf Science*, http://dx.doi.org/10.1016/j.ecss.2012.02.002.

Cossarini G, Solidoro C. (2008). Global sensitivity analysis of trophodynamic model of the Gulf of Trieste, *Ecological modelling*. 212: 16-27. doi:10.1016/j.ecolmodel.2007.10.009.

651 Costa FS, Pires MMDS, Nassar S M (2013). Analysis of Bayesian classifier
652 accuracy. *Journal of Computer Science, 9*(11): 1487-1495. doi:
653 10.3844/jcssp.2013.1487.1495.

654

655 Cozzi S, Ivancic I, Catalano G, Djakovac T, Degobbis D. (2004). Dynamics of the
656 oceanographic properties during mucilage appearance in the Northern Adriatic Sea: analysis
657 of the 1997 event in comparison to earlier events. *Journal of Marine Systems*, **50** (3-4): 223-
658 241. doi: 10.1016/j.jmarsys.2004.01.007.

659

660 Cozzi S, Falconi C, Comici C, Čermelj B, Kovač N, Turk V, Giani M. (2012). Recent
661 evolution of river discharges in the Gulf of Trieste and their potential response to climate
662 changes and anthropogenic pressure. *Estuarine, Coastal and Shelf Science*, 115: 14-24.

663

664 Danovaro R, Armeni M, Luna GM, Corinaldesi C, Dell'Anno A, Ferrari CR, Fiordelmondo
665 C, Gambi C, Gismondi M, Manini E, Mecoszzi M, Peronne FM, Pusceddu A, Giani M.
666 (2005). Exo-enzymatic activities and dissolved organic pools in relation to mucilage
667 development int eh Northern Adriatic Sea. *Science of the Total Environment*, 353(1-3): 189-
668 203. doi: 10.1016/j.scitotenv.2005.09.029.

669

670 DeLong E. F., Franks D. G., Alldredge A. L. 1993. Phylogenetic diversity of aggregate-
671 attached vs. free-living marine bacterial assemblages. Limnol Oceanogr, 38: 924-934

672

673 Del Negro P, Crevatin E, Larato C, Ferrari C, Totti C, Pompei M, Giani M. Berto M, Fonda
674 Umani S. (2005). Mucilage microcosms. *Science Total Environment*, 353:258–269. doi:
675 10.1016/j.scitotenv.2005.09.018.

676

677 Elskens, M, Baeyens, W, Brion, N, De Galan, S, Goeyens, L, De Brauwere, A. (2005).
678 Reliability of N flux rates estimated from 15 N enrichment and dilution experiments in
679 aquatic systems. *Global Biogeochemical Cycles*,**19**(4). doi: 10.1029/2004GB002332.

680

681 Faganeli J, Kovac N, Leskovšek H, Pezdič J. (1995). Sources and fluxes of particulate
682 organic matter in shallow coastal waters characterized by summer macroaggregate formation.
683 *Biogeochemistry*, 29:71–88.

684

685 Fajon C, Cauwet G, Lebaron P, Terzić S, Ahel M, Malej A, Mozetič P, Turk V. (1999). The
686 accumulation and release of polysaccharides by planktonic cells and the subsequent bacterial
687 response during a controlled experiment. *FEMS microbiology, ecology*, 29: 351-363.

688

689 Fawcett SE, Ward BB. (2011). Phytoplankton succession and nitrogen utilization during the
690 development of an upwelling bloom. *Marine Ecology Progress Series*, 428: 13-31.

691

692 Fehling J, Davidson K, Bolch CJS, Brand TD, Narayanaswamy BE. (2012). The relationship
693 between phytoplankton distribution and water column characteristics in North West European
694 shelf sea waters. *PLoS ONE*, 7(3): e34098. doi:10.1371/journal.pone.0034098.

695

696 Fletcher PJ, Kelble CR, Nuttle WK, Kiker GA. (2014). Using the integrated ecosystem
697 assessment framework to build consensus and transfer information to managers. *Ecological*
698 *Indicators*, doi: 10.1016/j.ecolind.2014.03.024.

699

700 Flores JM, Gámez JA, Martínez AM. (2014). Domains of competence of the semi-naive
701 Bayesian network classifiers. Information Sciences, 260: 120-148, ISSN 0020-0255. doi:
702 10.1016/j.ins.2013.10.007.

703

704 Fonda Umani S, Franco P, Ghirardelli E, Malej A. Outline of oceanography and the plankton
705 of the Adriatic Sea. In: Colombo, G., Ferrari, I., Ceccherelli, V. U., Rossi, R., (Eds.), Marine
706 Eutrophication and Population Dynamics. Proceedings of the 25th EMBS. Olsen & Olsen,
707 Fredensborg, 1992; pp. 347-365.

708

709 Fonda Umani S, Del Negro P, Larato C, De Vittor C, Cabrini M, Celio M, Falconi C,
710 Tamberlich F, Azam F. (2007). Major inter-annual variations in microbial dynamics in the
711 Gulf of Trieste (northern Adriatic Sea) and their ecosystem implications. *Aquatic Microbial*
712 *Ecology*, 46: 163–175.

713

714 Francé J, Mozetič P. (2012). Preliminary study on vertical migrations of dinoflagellates in a
715 dynamic coastal sea (Gulf of Trieste, northern Adriatic). *Acta Adriatica*, 53(2): 181-188.

716

717 Fuhrman JA, Hewson I, Schwalbach MS, Steele MS, Brown JA, Naeem S. (2006). Annually
718 reoccurring bacterial communities are predictable from ocean condi- tions. *Proc Natl Acad*
719 *Sci USA*, 103: 13104–13109.

720

721 Giani M, Degobbis D, Rinaldi A (2005) Mucilages in the Adriatic and Tyrrhenian Seas. Sci
722 Total Environ 353:1–2

723 Giani, M., Djakovac, T., Degobbis, D., Cozzi, S., Solidoro, C., and Fonda Umani, S. (2012)
724 Recent changes in the marine ecosystems of the northern Adriatic Sea. Estuar Coast Shelf Sci
725 115: 1–13.

726

727 Gilbert JA, Steele JA, Caporaso G, Steinbru L, Reeder J, Temperton B, Huse S, McHardy
728 AC, Knight R, Joint I, Somerfield P, Fuhrman JA, Field D. (2012). Defining seasonal marine
729 microbial community dynamics. *ISME Journal*, 6: 298–308.

730

731 Green EP, Dagg MJ. (1997). Mesozooplankton associations with medium to large marine
732 snow aggregates in the northern Gulf of Mexico. *Journal of Plankton Research*, **19**(4): 435-
733 447. doi: 10.1093/plankt/19.4.435.

734

735 Goericke, R. (2002). Top-down control of phytoplankton biomass and community structure
736 in the monsoonal Arabian Sea. *Limnology and Oceanography*, 47: 1307-1323.

737

738 Guillard RRL, Ryther JH. (1962). Studies of marine planktonic diatoms. I. *Cyclotella nana*
739 Hustedt and *Detonula confervacea* Cleve. *Canadian Journal of Microbiology*, 8: 229-239.

740

741 Hagström Å, Larsson U. (1984). Diel and seasonal variation in growth rates of pelagic
742 bacteria. *Plenum Publishing Corporation*, 249-262.

743

744 Hamilton, S, Pollino, C & Jakeman, A 2015, 'Habitat suitability modelling of rare species
745 using Bayesian networks: Model evaluation under limited data', Ecological Modelling, vol.
746 299, pp. 64-78. Doi: 10.1016/j.ecolmodel.2014.12.004.

747

748 Hurwitz BL, Westveld AH, Brum JR, Sullivan MB. (2014). Modeling ecological drivers in
749 marine viral communities using comparative metagenomics and network analyses.
750 Proceedings of the National Academy of Sciences of the United States of America, Vol.111:
751 10714-9. doi: 10.1073/pnas.1319778111.

752

753 Isci S, Dogan H, Ozturk C, Otu HH. Bayesian network prior: network analysis of biological
754 data using external knowledge. *Bioinformatics*. 2014;30(6):860-867.
755 doi:10.1093/bioinformatics/btt643.

756

Ivančić I, Radić T, Lyons DM, Fuks D, Precali R, Kraus R. (2009). Alkaline phosphatase activity in relation to nutrient status in the northern Adriatic Sea. *Marine Ecology Progress Series*, 378: 27-35. doi: 10.3354/meps07851

Kriest I. (2002). Different parameterizations of marine snow in a 1D-model and their influence on representation of marine snow, nitrogen budget and sedimentation. *Deep-Sea Research Part I*, **49**: 2133-2162. doi: 10.1016/S0967-0637(02)00127-9.

Krom MD, Kress N, Brenner S. (1991). Phosphorus limitation of primary productivity in the easter Mediterranean Sea. *Limnology Oceanography*, 36:424-432.

Kocabas V, Dragicevic S, McCann E. (2012). Integration of a GIS-Bayesian Network agent-based model in a planning support system as framework for policy generation. *URISA Journal*, **24**(1): 35-52.

Larsen PE, Field D, Gilbert JA. (2012). Predicting bacterial community assemblages using an artificial neural network approach. *Nature methods*, 9(6): 621-625.

Lee HS, Lee JH. (1995). Continuous monitoring of short term dissolved oxygen and algal dynamics. *Water Research*, 29(12): 2789-2796. doi: 10.1016/0043-1354(95)00126-6.

Levinton JS. Marine Biology, Function, Biodiversity, Ecology. Fourth Edition, Jeffrey S. Levinton July 2013. SBN: 9780199857128.Levontin P, Kulmala S, Haapasaari P, Kuikka S. (2011). Integration of biological, economic, and sociological knowledge by Bayesian belief networks: the interdisciplinary evaluation of potential management plans for Baltic salmon. *Journal of Marine Science*, 68(3): 632-638. doi: 10.1093/icesjms/fsr004.

784 Li C, Li H, (2013). Bayesian network classifiers for probability-based metrics. *Journal of*
785 *Experimental & Theoretical Artificial Intelligence*, 25(4). doi:
786 10.1080/0952813X.2013.782345.

787

788 Lindeman R. (1942). The trophic-dynamic aspect of ecology. *Ecology*, 23: 399-418.

789

790 Lowe CD, Gilbert AJ, Mee LD. (2014). Human-environment interaction in the Baltic Sea.
791 *Marine Policy*, 43: 46-54. doi: 10.1016/j.marpol.2013.03.006.

792

793 Liu, Feng ; Tian, Fengzhan ; Zhu, Qiliang (2007). A Novel Ordering-Based Greedy Bayesian
794 Network Learning Algorithm on Limited Data. Seventh IEEE International Conference on
795 Data Mining Workshops, pp.495-500. doi: 10.1109/ICDMW.2007.13.

796

797 Lucena-Moya P, Brawata R, Kath J, Harrison E, El Sawah S, Dyer F, Discretization of
798 continuous predictor variables in Bayesian networks: An ecological threshold approach,
799 Environmental Modelling & Software, Volume 66, April 2015, Pages 36-45, ISSN 1364-
800 8152, http://dx.doi.org/10.1016/j.envsoft.2014.12.019.

801

802 Malačič V, Celio M, Čermelj B, Bussani A, Comici C. (2006). Interannual evolution of
803 seasonal thermohaline properties in the Gulf of Trieste (northern Adriatic) 1991-2003.
804 *Journal of Geophysical Research*, 111: C08009. doi:10.1029/2005JC003267.

805

806 Malej A, Mozetič P, Malačič V, Terzić S, Ahel M. (1995). Phytoplankton response to
807 freshwater inputs in a small semi-enclosed gulf (Gulf of Trieste, Adriatic Sea). *Marine*
808 *Ecology Progress Series*, 120: 111-121.

809

Malej A, Mozetič P, Malačič V, Turk V. (1997). Response of summer phytoplankton to episodic meteorological events (Gulf of Trieste, Adriatic Sea). *Marine Ecology*, 18: 273-28

Malfatti F, Turk V, Tinta T, Mozetic P, Manganelli M, Samo TJ, Ugalde JA, Kovac N, Stefanelli M, Antonioli M, Fonda Umani S, Del Negro P, Cataletto B, Hozic A, Ivosevic DeNardis N, Zutic V, Svetlicic V, Misic Radic T, Radic T, Fuks D, Azam F. (2014). Microbial mechanisms coupling carbon and phosphorus cycles in phosphorus-limited northern Adriatic Sea. *Science of the Total Environment*, **470-471**: 1173-1183. doi: 10.1016/j.scitotenv.2013.10.040.

Mantyka- Pringle CS, Martin TG, Moffat DB, Linke S, Rhodes JR. (2014). Understanding and predicting the combined effects of climate change and land- use change on freshwater macroinvertebrates and fish. *Journal of Applied Ecology*, **51**(3): 572-581. doi: 10.1111/1365-2664.12236.

McDonald KS, Ryder DS, Tighe M. (2015). Developing best-practice Bayesian Belief Networks in ecological risk assessments for freshwater and estuarine ecosystems: A quantitative review. *Journal of Environmental Management*, 150: 190-200. doi: 10.1016/j.jenvman.2015.02.031.

McDonald KS, Tighe M, Ryder DS. (2016) An ecological risk assessment for managing and predicting trophic shifts in estuarine ecosystems using a Bayesian network. *Environmental Modelling & Software*, 85: 202-216. doi: 10.1016/j.envsoft.2016.08.014.

Melbourne-Thomas J, Wotherspoon S, Raymond B, Constable A. (2012). Comprehensive evaluation of model uncertainty in qualitative network analyses. *Ecological Monographs*, **82**: 505–519. doi:10.1890/12-0207.1.

838 Metcalf S, Putten E, Frusher S, Tull M, Marshall N. (2014). Adaptation options for marine
839 industries and coastal communities using community structure and dynamics. *Sustainability*
840 *Science*, **9**(3): 247-261. doi: 10.1007/s11625-013-0239-z.

841

842 Miller CB. Biological Oceanography. Blackwell Publishing, 2004; 402 p.

843

844 Monticelli L, Caruso G, Decembrini F, Caroppo C, Fiesoletti F. (2014). Role of Prokaryotic
845 Biomasses and Activities in Carbon and Phosphorus Cycles at a Coastal, Thermohaline Front
846 and in Offshore Waters (Gulf of Manfredonia, Southern Adriatic Sea). *Microbial Ecology*,
847 67(3): 501-519. doi: 10.1007/s00248-013-0350-9.

848

849 Moran SB, Lomas MW, Kelly RP, Gradinger R, Iken K, Mathis JT. (2012). Seasonal
850 succession of net primary productivity, particulate organic carbon export, and autotrophic
851 community composition in the eastern Bering Sea. *Deep-Sea Research Part II*, **65-70**: 84-97.
852 doi: 10.1016/j.dsr2.2012.02.011.

853

854 Mozetič P, Solidoro C, Cossarini G, Socal G, Precali R, Francé J, Bianchi F, De Vittor C,
855 Smodlaka N, Fonda Umani S. (2010). Recent trends towards oligotrophication of the
856 Northern Adriatic: Evidence from Chlorophyll a time series. *Estuaries and Coasts*, 33: 362-
857 375.

858

859 Mozetič P, Francé J, Kogovšek T, Talaber I, Malej A. (2012). Plankton trends and
860 community changes in a coastal sea (northern Adriatic): Bottom-up vs. top-down control in
861 relation to environmental drivers. *Estuarine, Coastal and Shelf Science*, 115: 138-148.

862

863 Najdek M, Degobbis D, Miokovic c D, Ivancic I. (2002). Fatty acid and phytoplankton
864 compositions of different types of mucilaginous aggregates in the northern Adriatic. *J*
865 *Plankton Res* 24:429–441.

Norsys Software Corporation, Netica Version 4.16, 2010. www.norsys.com

Nojavan AF, Qian SS, Paerl HW, Reckhow KH, Albright EA. (2014). A study of anthropogenic and climatic disturbance of the New River Estuary using a Bayesian Belief Network. *Marine Pollution Bulletin*. doi: 10.1016/j.marpolbul.2014.04.011.

O'Connor DJ, Di Toro DM. (1970). Photosynthesis and oxygen balance in streams. *Journal of the Sanitary Engineering Division*, 96(2): 547-571.

Oddo P, Pinardi N, Zavatarelli M, A numerical study of the interannual variability of the Adriatic Sea (2000–2002), Science of The Total Environment, Volume 353, Issues 1–3, 15 December 2005, Pages 39-56, ISSN 0048-9697. doi: 10.1016/j.scitotenv.2005.09.061.

Perlinski, AT, Paige, GB, Mcclaran, MP. (2014). Evaluating a State-and-Transition Model Using a Long-Term Dataset. *Rangeland Ecology & Management*, **67**(2): 173-182. doi: 10.2111/REM-D-12-00036.1.

Pollino, CA, Woodberry, O, Nicholson, A, Korb, K, Hart, BT. (2007). Parameterisation and evaluation of a Bayesian network for use in an ecological risk assessment. *Environmental Modelling and Software*, 22(8): 1140-1152. doi: 10.1016/j.envsoft.2006.03.006.

Pomeroy LR, Wiebe WJ. (2001). Temperature and substrates as interactive limiting factors for marine heterotrophic bacteria. *Aquatic Microbial Ecology*, **23**(2): 187-204.

Puddu A, La Ferla R, Allegra A, Bacci C, Lopez M, Oliva F, Pierotti C. (1997). Seasonal and spatial distribution of bacterial production and biomass along a salinity gradient (Northern Adriatic Sea). *Hydrobiologia*, 363(1): 271-282 . doi: 10.1023/A:1003169620843.

Ribera d'Alcalá M, Conversano F, Corato F, Licandro P, Mangoni O, Marino D, Mazzocchi MG, Modigh M, Montresor M, Nardella M, Saggiomo V, Sarno D, Zingone A. (2004). Seasonal patterns in plankton communities in a pluriannual time series at a coastal Mediterranean site (Gulf of Naples): an attempt to discern recurrences and trends. *Scientia Marina*, 68(Suppl. 1): 65-83.

Scholten L, Scheidegger A, Reichert P, Maurer M, Combining expert knowledge and local data for improved service life modeling of water supply networks, Environmental Modelling & Software, Volume 42, April 2013, Pages 1-16, ISSN 1364-8152, http://dx.doi.org/10.1016/j.envsoft.2012.11.013.

Signell RP, Carniel S, Cavaleri L, Chiggiato J, Doyle JD, Pullen J, Sclavo M. (2005). Assessment of wind quality for oceanographic modelling in semi-enclosed basins. *Journal of Marine Systems*, **53**(1): 217-233. doi: 10.1016/j.jmarsys.2004.03.006.

Smayda TJ. (1997). Harmful algal blooms: Their ecophysiology and general relevance to phytoplankton blooms in the sea. *Limnology and Oceanography*, 42(5-2): 1137-1153.

Smith, D.C., Steward, G.F., Long, R.A., Azam, F. (1995). Bacterial Mediation of Carbon Fluxes During a Diatom Bloom in a Mesocosm. Deep-Sea Research Part II-Topical Studies in Oceanography, 42 (1), 75-97.

917 Stelzenmüller V, Lee J, Garnacho E, Rogers SI. (2014). Assessment of a Bayesian Belief
918 Network-GIS framework as a practical tool to support marine planning. *Marine Pollution*
919 *Bulletin*, **60**(10): 1743-1754. doi: 10.1016/j.marpolbul.2010.06.024.

920

921 Talaber I, Francé J, Mozetič P. (2014). How phytoplankton physiology and community
922 structure adjust to physical forcing in a coastal ecosystem (northern Adriatic Sea).
923 *Phycologia*, 53(1): 74-85. doi: 10.2216/13-196.1.

924

925 Tamše S, Mozetič P, Francé J, Ogrinc N. (2014). Stable isotopes as a tool for nitrogen source
926 identification and cycling in the Gulf of Trieste (Northern Adriatic). *Continental Shelf*
927 *Research*, 91: 145-157. doi: 10.1016/j.csr.2014.09.009.

928

929 Taylor AG, Landry MR, Selph KE, Wokuluk JJ. (2014). Temporal and spatial patterns of
930 microbial community biomass and composition in the Southern California Current
931 Ecosystem. *Deep-Sea Research Part II*. doi: 10.1016/j.dsr2.2014.02.006.

932

933 Teeling, H., Fuchs, M.B., Becher, D., Klockow, C., Gardebrecht, A., Behnke, C.M., *et al*.
934 (2012) Substratecontrolled succession of marine bacterioplankton populations induced by a
935 phytoplankton bloom. *Science* 336: 608–611.

936

937 Tinta T, Vojvoda J, Mozetič P, Talaber I, Vodopivec M, Malfatti F, Turk V. (2015).
938 Bacterial community shift is induced by dynamic environmental parameters in a changing
939 coastal ecosystem (northern Adriatic, NE Mediterranean Sea) - a 2 year time series study.
940 *Environmental microbiology*, 17(10): 3581-3596. doi:10.1111/1462-2920.12519.

941

942 Turk V, Å Hagström, N Kovač, J Faganeli. (2010). Composition and function of mucilage
943 macroaggregates in the northern Adriatic. V: SAME 11 - The 11th Symposium on Aquatic
944 Microbial Ecology, August 30 - September 04 2009, Piran, Slovenia. DEL GIORGIO, Paul

A. (ur.). *Progress and perspectives in aquatic microbial ecology*, Aquatic microbial ecology, 61, 279-289. doi: 10.3354/ame01447.

Turner, J T. (2014) Zooplankton fecal pellets, marine snow, phytodetritus and the ocean's biological pump. *Progress in Oceanography*, 130: 205-248.

Van Dam, R, Humphrey, C, Harford, A, Sinclair, A, Jones, D, Davies, S, Storey, A. (2014). Site-specific water quality guidelines: 1. Derivation approaches based on physicochemical, ecotoxicological and ecological data. *Environmental Science and Pollution Research International*, **21**(1): 118-30. doi: 10.1007/s11356-013-1780-0.

Van Grieken M, Lynam T, Coggan A, Whitten S, Kroon F. (2013). Cost effectiveness of design-based water quality improvement regulations in the Great Barrier Reef Catchments. *Agriculture, Ecosystems and Environment*, **180:** 157-165. doi: 10.1016/j.agee.2011.08.010.

Verdugo P, Alldredge AL, Azam F, Kirchman DL, Passow U, Santsch PH. (2004). The oceanic gel phase: a bridge in the DOM–POM continuum. *Marine Chemistry*, 92: 67– 85.

Weitz JS, Stock CA, Wilhelm SW, Bourouiba L, Buchan A, Coleman ML, Follows MJ, Fuhrman JA, Jover LF, Lennon JT, Middelboe M, Sonderegger DL, Suttle CA, Taylor BP, Thingstad TF, Wilson WH, Wommack KE. (2015). A multitrophic model to quantify the effects of marine viruses on microbial food webs and ecosystem processes. *ISME Journal*, 9:1352-1364. doi:10.1038/ismej.2014.220

Wikner J, Hagström Å. (1991). Annual study of bacterioplankton community dynamics. *Limnology Oceanography*, 36:1313-1324.

Wright PA. (1995). Nitrogen excretion: three end products, many physiological roles. *Journal of Experimental Biology*, 198: 273-281.

Xu J, Ma B. (2014). Study of network public opinion classification method based on naive bayesian algorithm in hadoop environment. *Applied Mechanics and Materials, 519-520*: 58-61. doi: 10.4028/www.scientific.net/AMM.519-520.58.

Yunev OA, Carstensen J, Moncheva S, Khaliulin A, Aertebjerg G, Nixon S. (2007). Nutrient and phytoplankton trends on the western Black Sea shelf in response to cultural oligotrophication and climate change. *Estuarine, Coastal and Shelf Science*, 74: 63-76. doi:10.1016/j.ecss.2007.03.030

Supporting Information 1: Node configuration

Table S1.1 The definition and scientific rationale for each node in the network.

Node	Definition	States	Ecological Importance	Cases
Data	The source of the data used.	Field Observations Experimental	Experimental ecosystems can produce unreliable results (Elskens et al. 2005; Brock et al., 2015)	129
Season	Calendar seasons	Summer Autumn Winter Spring	The seasonal variability of physical and chemical mechanisms in the system can influence the microbial processes occurring in the ecosystem (Tinta et al., 2015).	129
Depth (m)	Probabilities extrapolated from Tinta et al. (2015).	Surface (5) Bottom (20)	Vertical distribution of physical-chemical parameters that influence the growth and abundance of plankton organisms (e.g., Fehling et al., 2012).	129
Water Temperature (ºC)	States derived from <10, 10-25, 25- 75, 75-90, >90 percentiles of data from Tinta et al. (2015).	Minimum 0 - 9.78 Mild 9.78- 11.73 Mean 11.73 – 19.69 Moderate 19.69 -23.03 Maximum >23.03	Temperature influences the rate of biological processes (Pomeroy and Wiebe, 2001).	100
Salinity	States derived from <10, 10-25, 25- 75, 75-90, >90 percentiles of data from Tinta et al. (2015).	Minimum 0 – 34.52 Mild 34.52 – 36.13 Mean 36.13 – 37.57 Moderate 37.57 – 37.69	The salinity governs physical, chemical and biological processes (Levinton, 2013).	100

		Maximum >37.69		
Dissolved oxygen (mg/L)	States derived from <10, 10-25, 25- 75, 75-90, >90 percentiles of data from Tinta et al. (2015).	Minimum 0 - 4.03 Mild 4.03 - 4.80 Mean 4.80 - 5.80 Moderate 5.80 - 6.45 Maximum >6.45	The dissolved oxygen concentration is the one of the major factor that determines the type and abundance of organisms as well as biochemical processes (O'Connor and Di Toro, 1970; Lee and Lee, 1995).	90
Total Nitrogen (TN) (µmol/L)	States derived from <10, 10-25, 25- 75, 75-90, >90 percentiles of data from Tinta et al. (2015) and Malfatti et al. (2014).	Minimum 0-9.07 Mild 9.07-11.63 Mean 11.63-23.50 Moderate 23.50- 26.31 Maximum >26.31	Giani et al. (2012) reported POC aggregate accumulation has a hyperbolic relationship to TN concentrations.	127
Total Phosphorus (TP) (µmol/L)	States derived from <10, 10-25, 25- 75, 75-90, >90 percentiles of data from Tinta et al. (2015) and Malfatti et al. (2014).	Minimum 0-0.19 Mild 0.24-0.19 Mean 0.24-0.35 Moderate 0.35-0.4 Maximum >0.4	Giani et al. (2012) reported POC aggregate accumulation has a linear relationship to TP concentrations.	127
Ammonium (NH_4^+) (µmol/L)	States derived from <10, 10-25, 25- 75, 75-90, >90 percentiles of data outlined in Tinta et al. (2015) and Malfatti et al. (2014).	Minimum 0-0.20 Mild 0.20-0.32 Mean 0.32-0.98 Moderate 0.98-1.57 Maximum >1.57	A by-product of OM degradation that increases in concentrations in the water column through release from sediment, excretion of zooplankton (Wright, 1995) and inputs from land (e.g., sewage) (Brigolin et al., 2011). It is a bioavailable form of N for the biota and can serve as an energy source for bacteria (Miller,	127

			2004).	
Nitrite (NO_2^-) (µmol/L)	States derived from <10, 10-25, 25- 75, 75-90, >90 percentiles of data outlined in Tinta et al. (2015) and Malfatti et al. (2014).	Minimum 0-0.04 Mild 0.04-0.09 Mean 0.09-0.46 Moderate 0.46-0.88 Maximum >0.88	A bioavailable form of N that is required for photosynthetic processes and microbial processes (Miller, 2004).	127
Nitrate (NO_3^-) (µmol/L)	States derived from <10, 10-25, 25- 75, 75-90, >90 percentiles of data outlined in Tinta et al. (2015) and Malfatti et al. (2014).	Minimum 0-0.45 Mild 0.45-0.83 Mean 0.83-4.39 Moderate 4.39-5.86 Maximum >5.86	A bioavailable form of N that is required for photosynthetic processes and microbial processes (Miller, 2004).	125
Silicate SiO_4^{4-} (µmol/L)	States derived from <10, 10-25, 25- 75, 75-90, >90 percentiles of data outlined in Tinta et al. (2015) and Malfatti et al. (2014)	Minimum 0-2.69 Mild 2.69-5.44 Mean 5.44-7.31 Moderate 7.31-9.35 Maximum >9.35	Silicon (in the form of orthosilicate ion) is a major nutrient for diatoms and silicoflagellates (Miller, 2004).	127
Orthophosphate (PO_4^{3-}) (µmol/L)	States derived from <10, 10-25, 25- 75, 75-90, >90 percentiles of data outlined in Tinta et al. (2015) and Malfatti et al. (2014).	Minimum 0-0.04 Mild 0.04-0.05 Mean 0.05-0.1 Moderate 0.1 – 0.15 Maximum 0.15- 5	A bioavailable form of P that is required for photosynthetic processes and microbial processes (Krom et al., 1991).	127
Dissolved Organic Phosphorus (DOP)	States derived from <25, 25-75, >75 percentiles of data outlined	Minimum 0-0.11 Mean 0.11-0.18	Organic phosphorus pool that is cleaved by the alkaline phosphatase enzyme present in bacteria	19

(µmol/L)	in Malfatti et al. (2014).	Maximum >0.18	and phytoplankton (Ivančić et al. 2009)	
Particulate Organic Carbon (POC) (µmol/L)	Percentiles of data outlined in Malfatti et al. (2014).	Minimum 0 - 9.82 Mean 9.82 - 68.33 Maximum >68.33	The origin and variation in chemical composition vary annually and can be affected by type of microbial community provoking blooms (Faganeli et al., 1995)	16
Dissolved Organic Carbon (DOC) (µmol/L)	States derived from <25, 25-75, >75 percentiles of data outlined in Malfatti et al. (2014).	Minimum 0-91.29 Mean 91.29-160.23 Maximum >160.23	During algal growth substantial amount of DOC could be released and subsequently utilised by heterotrophic bacteria and influence the accumulation (Azam et al., 1983).	20
Chlorophyll *a* (µg/L) (Chl *a*)	States derived from <10, 10-25, 25- 75, 75-90, >90 percentiles of data from Tinta et al. (2015) and Malfatti et al. (2014).	Minimum 0 - 0.31 Low 0.31 - 0.42 Mean 0.42 - 0.99 High 0.99 - 1.62 Maximum >1.62	The concentration of chlorophyll *a* indicates the biomass of phytoplankton, i.e. phytoplankton stock in the water column and is the key light-absorbing pigment involved in photosynthesis (Miller, 2004).	124
Total Phytoplankton abundance (cells/L)	States derived from <25, 25- 75, 75-90, >90 percentiles of data outlined in Tinta et al. (2015) and Malfatti et al. (2014).	Low 0 - 300000 Mean 300000 - 600000 High 600000 - 900000 Maximum >900000	The amount of photosynthesised carbon can strongly vary with changes in the phytoplankton abundance and composition (Fonda Umani et al., 2005).	95
Net Primary Production (PP) (µg C/L h)	States derived from <25, 25-75, >75 percentiles of data outlined in and Malfatti et al. (2014).	Minimum 0 - 4.07 Mean 4.07 - 16.33 Maximum >16.33	Net PP indicates the amount of inorganic C fixed into autotrophic biomass via photosynthetic processes within a specified time period and is subsequently available to higher trophic levels (Lindeman, 1942).	17

Bacterial Carbon Production (µgC/L day)	States derived from <10, 10-25, 25- 75, 75-90, >90 percentiles of data outlined in Tinta et al. (2015) and Malfatti et al. (2014).	Minimum 0 - 1.07 Low 1.07 - 2.1 Mean 2.1 - 8.75 High 8.75 - 11.44 Maximum >11.44	Bacterial growth rate is dependent on nutrient availability and on temperature (Hagström and Larsson, 1984).	99
Bacteria Abundance (cells x 10^8/L)	States derived from <10, 10-25, 25- 75, 75-90, >90 percentiles of data outlined in Tinta et al. (2015) and Malfatti et al. (2014).	Minimum < 1.45 Low 1.45 -2.06 Mean 2.06 -5.71 High 5.71- 9.57 Maximum > 9.57	Bacteria abundance varies seasonally and depends to a large extend on vertical physical processes and nutrient concentrations (Wikner and Hagström, 1991).	109
Alkaline phosphatase (nM/h)	States derived from percentiles of data outlined in Malfatti et al. (2014).	Minimum 0 - 24.28 Mean 24.28 - 325.52 Maximum > 325.52	Enzyme that hydrolyses phosphate from phosphorus rich compounds (Celussi and Del Negro, 2011).	14
Aminopeptidase (nM/h)	States derived from percentiles of data outlined in Malfatti et al. (2014).	Minimum 0 - 80.48 Mean 80.48 - 466.90 Maximum > 466.90	Enzyme that hydrolyses aminoacids from proteins (Celussi and Del Negro, 2011).	14
Dominant phytoplankton	States derived from percentiles of data outlined in Tinta et al. (2015) and Malfatti et al. (2014).	Nanoflagellates Diatoms Dinoflagellates Unidentified	Seasonal shifts in phytoplankton community composition that influences autotrophic C availability e.g., a diatom or nanoflagellate dominated community structure (Mozetič et al., 2012; Moran et al., 2012; Taylor et al., 2014).	95
Nanoflagellates (cells/L)	States derived from <25, 25-75, 75-90, >90 percentiles of data	Low 0 - 100000 Mean 100000 - 500000	The abundance of each phytoplankton class/group reflects growth rates of that group,	95

	outlined in Tinta et al. (2015) and Malfatti et al. (2014).	High 500000 - 900000 Maximum >900000	but also physical processes (advection, horizontal mixing) that can promote different types of phytoplankton (Miller, 2004).	
Diatoms (cells/L)	States derived from <25, 25-75, 75-90, >90 percentiles of data outlined in Tinta et al. (2015) and Malfatti et al. (2014).	Low 0 - 4500 Mean 4500 - 174875 High 174875 - 487556 Maximum >487556		95
Dinoflagellates (cells/L)	States derived from <25, 25-75, 75-90, >90 percentiles of data outlined in Tinta et al. (2015) and Malfatti et al. (2014).	Low 0 - 6000 Mean 6000 - 37628.50 High 37628.50 - 52900 Maximum >52900		95
Coccolithophorids (cells/L)	States derived from <25, 25-75, 75-90, >90 percentiles of data outlined in Tinta et al. (2015) and Malfatti et al. (2014).	Low 0 - 6650 Mean 6650 - 42750 High 42750 - 69900 Maximum >69900		95
Silicoflagellates (cells/L)	States derived from <75, >75 percentiles of data outlined in Tinta et al. (2015) and Malfatti et al. (2014).	Mean 0 - 1000 High >1000		95
Non identified algae (cells/L)	States derived from <25, 25-75, 75-90, >90 percentiles of data from Tinta et al. (2015) and Malfatti et al. (2014).	Low 0 - 4000 Mean 4000 - 8000 High 8000 - 11900 Maximum >11900		95
Alphaproteobacteria	States derived from percentiles	Minimum 0 - 0.27	Variable environmental parameters (biotic and	16

(Relative abundance)	of data outlined in Tinta et al. (2015) and Malfatti et al. (2014).	Mean 0.27 - 0.37 Maximum >0.37	abiotic) affect bacterial community composition (Fuhrman et al., 2006; Gilbert et al., 2012). Different phylotypes of bacteria have diverse metabolism that influence the carbon degradation and accumulation processes (Fuhrman et al., 2006; Teeling et al., 2012). Sequence taxonomic identities (at > 97% similarity) were assigned using the genome Basic Local Alignment Search Tool (BLAST) at the National Center for Biotechnology Information (NCBI). Classification was done down to the bacterial family level. In order to take into account the libraries with different sequencing depths we expressed the contributions of distinct bacterial families as a percentage of the total number of sequences in each library (relative abundance) (Tinta et al., 2015).	
Rhodospirillaceae and Rhodobacteraceae (Relative abundance)	States derived from <75, >75 percentiles of data outlined in Tinta et al. (2015) and Malfatti et al. (2014).	Mean 0 - 0.143 High >0.143		16
Gamma-Proteobacteria (Relative abundance)	States derived from <25, 25-75, >75 percentiles of data outlined in Tinta et al. (2015) and Malfatti et al. (2014).	Low 0 – 0.149 Mean 0.149 - 0.181 High >0.181		16
Alteromonadaceae (Relative abundance)	States derived from percentiles of data outlined in Tinta et al. (2015) and Malfatti et al. (2014).	Mean 0 - 0.39 Maximum >0.39		16
SAR11 (Relative abundance)	States derived from percentiles of data outlined in Tinta et al. (2015) and Malfatti et al. (2014).	Minimum 0 – 0.28 Mean 0.28 - 0.68 Maximum > 0.68		16
SAR86 (Relative abundance)	States derived from percentiles of data outlined in Tinta et al. (2015) and Malfatti et al. (2014).	Minimum 0 – 0. 0.16 Mean 0.16 – 0.53 Maximum >0.53		16
Betaproteobacteria (Relative abundance)	States derived from <90, >90%percentiles of data outlined in Tinta et al. (2015) and	Mean 0 – 0.02 Maximum >0.02		16

	Malfatti et al. (2014).		
Deltaproteobacteria (Relative abundance)	States derived from <90, >90%percentiles of data outlined in Tinta et al. (2015) and Malfatti et al. (2014).	Mean 0 - 0.003 Maximum > 0.003	16
Epsilonproteobacteria (Relative abundance)	States derived from percentiles of data outlined in Tinta et al. (2015) and Malfatti et al. (2014).	Present Absent	16
Unclassified Proteobacteria (Relative abundance)	States derived from <90, >90%percentiles of data outlined in Tinta et al. (2015) and Malfatti et al. (2014).	Mean 0 - 0.003 Maximum > 0.003	16
Flavobacteria (Relative abundance)	States derived from percentiles of data outlined in Tinta et al. (2015) and Malfatti et al. (2014).	Minimum 0 – 0.07 Mean 0.07 - 0.23 Maximum >0.23	16
Sphingobacteria (Relative abundance)	States derived from percentiles of data outlined in Tinta et al. (2015).	Mean 0 - 0.05 Maximum >0.05	16
Cytophaga (Relative abundance)	States derived from percentiles of data outlined in Tinta et al. (2015) and Malfatti et al. (2014).	Mean 0 – 0.0001 Maximum >0.0001	16
Unclassified Bacteroidetes	States derived from <90, >90%percentiles of data outlined	Mean 0 - 0.003 Maximum >0.003	23

(Relative abundance)	in Tinta et al. (2015) and Malfatti et al. (2014).			
Actinobacteria (Relative abundance)	States derived from <90, >90%percentiles of data outlined in Tinta et al. (2015) and Malfatti et al. (2014).	Mean 0 – 0.05 Maximum >0.05		16
Cyanobacteria (Relative abundance)	States derived from percentiles of data outlined in Tinta et al. (2015) and Malfatti et al. (2014).	Minimum 0-0.07 Mean 0.07 - 0.26 Maximum >0.26		23
Planctomycetes (Relative abundance)	States derived from <90, >90%percentiles of data outlined in Tinta et al. (2015) and Malfatti et al. (2014).	Mean 0 - 0.04 Maximum >0.04		23
Verrumcomicrobia (Relative abundance)	States derived from percentiles of data outlined in Tinta et al. (2015) and Malfatti et al. (2014).	Mean 0 - 0.04 Maximum >0.04		23
Deferribacteria (Relative abundance)	States derived from <90, >90%percentiles of data outlined in Tinta et al. (2015) and Malfatti et al. (2014).	Absent Present		16
Dominant Bacteria (Relative abundance)	States derived from percentiles of data outlined in Tinta et al. (2015) and Malfatti et al. (2014).	SAR11 Proteobacteria	The bacteria present in the highest abundance at each observation point.	23

	Rhodobacteria
	SAR86
	Alteromonsdaceae

988

Supporting Information 2 Appendix of the scenario tests conducted in this study.

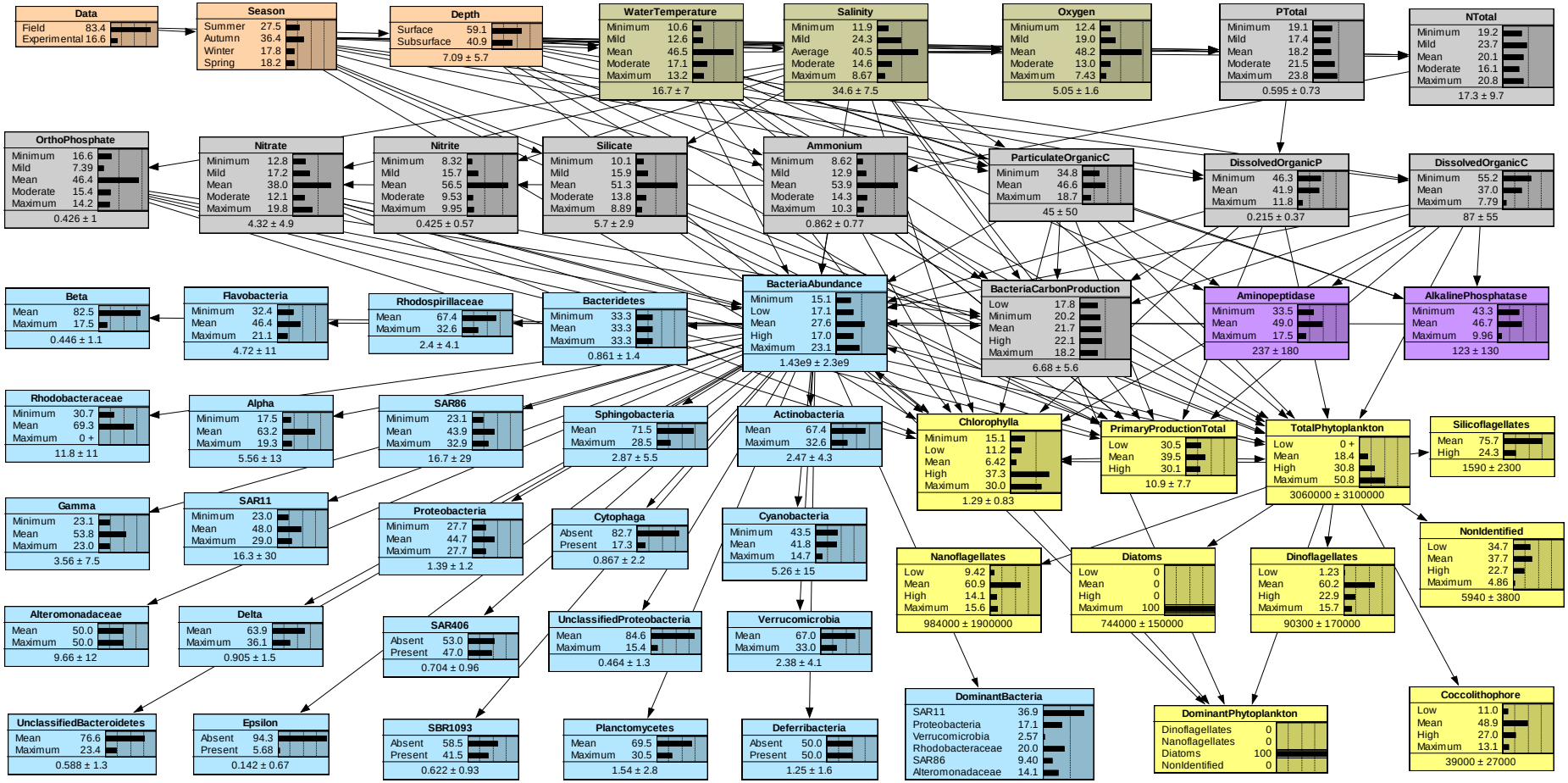
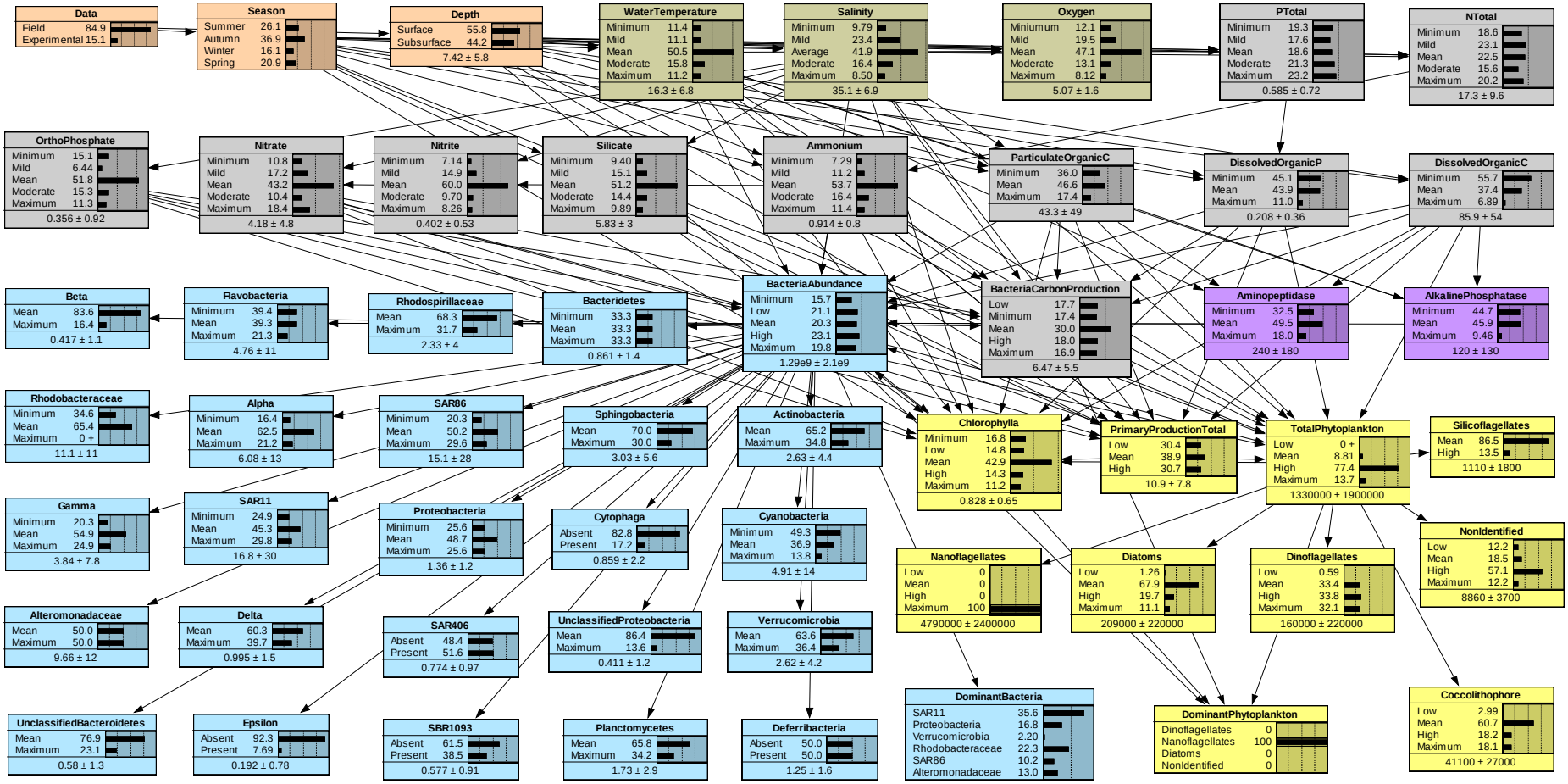


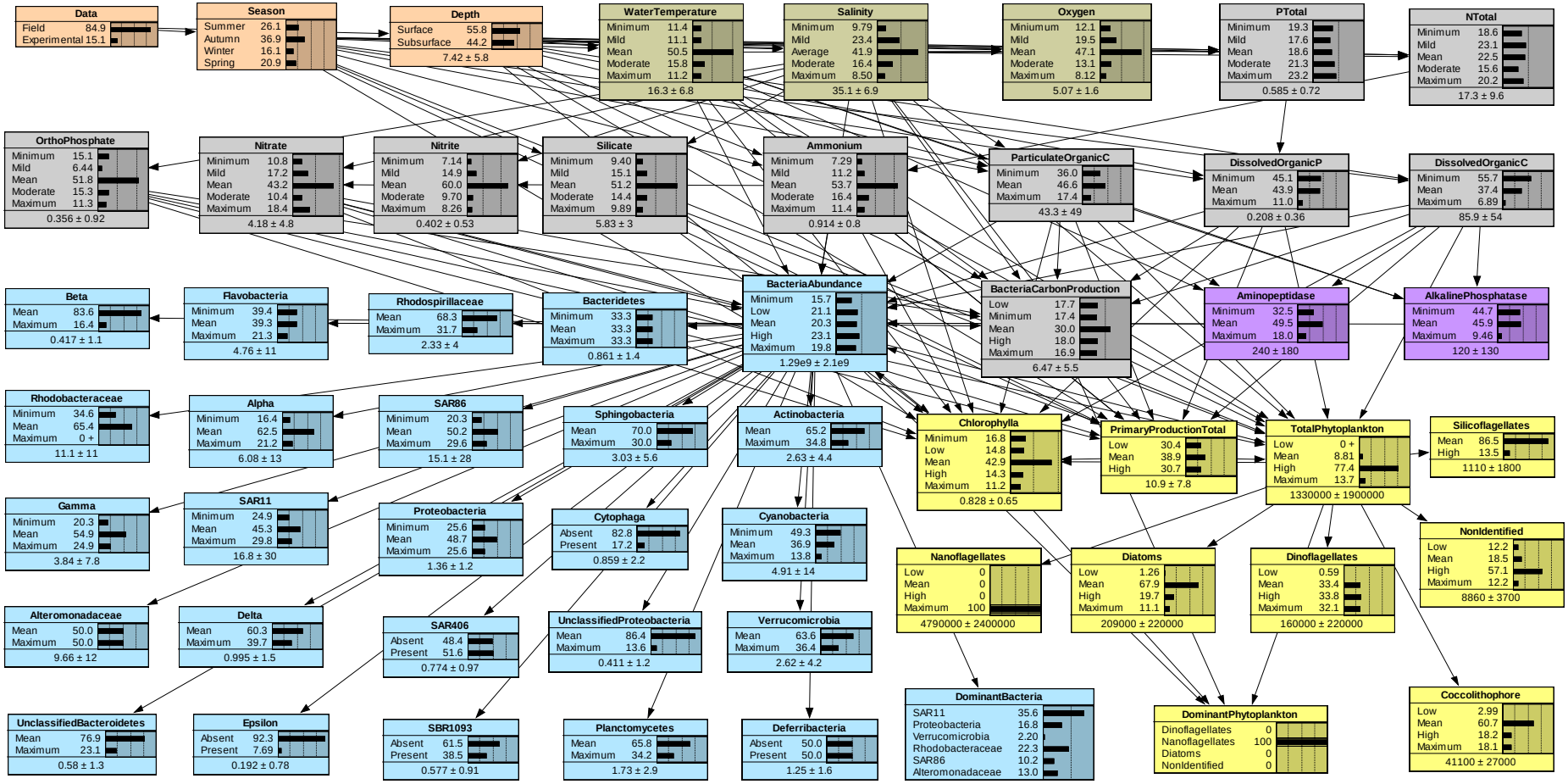
Fig S2.1 Scenario test for a diatom bloom.



992

993 **Fig. S2.2** Scenario test for a dinoflagellate bloom.

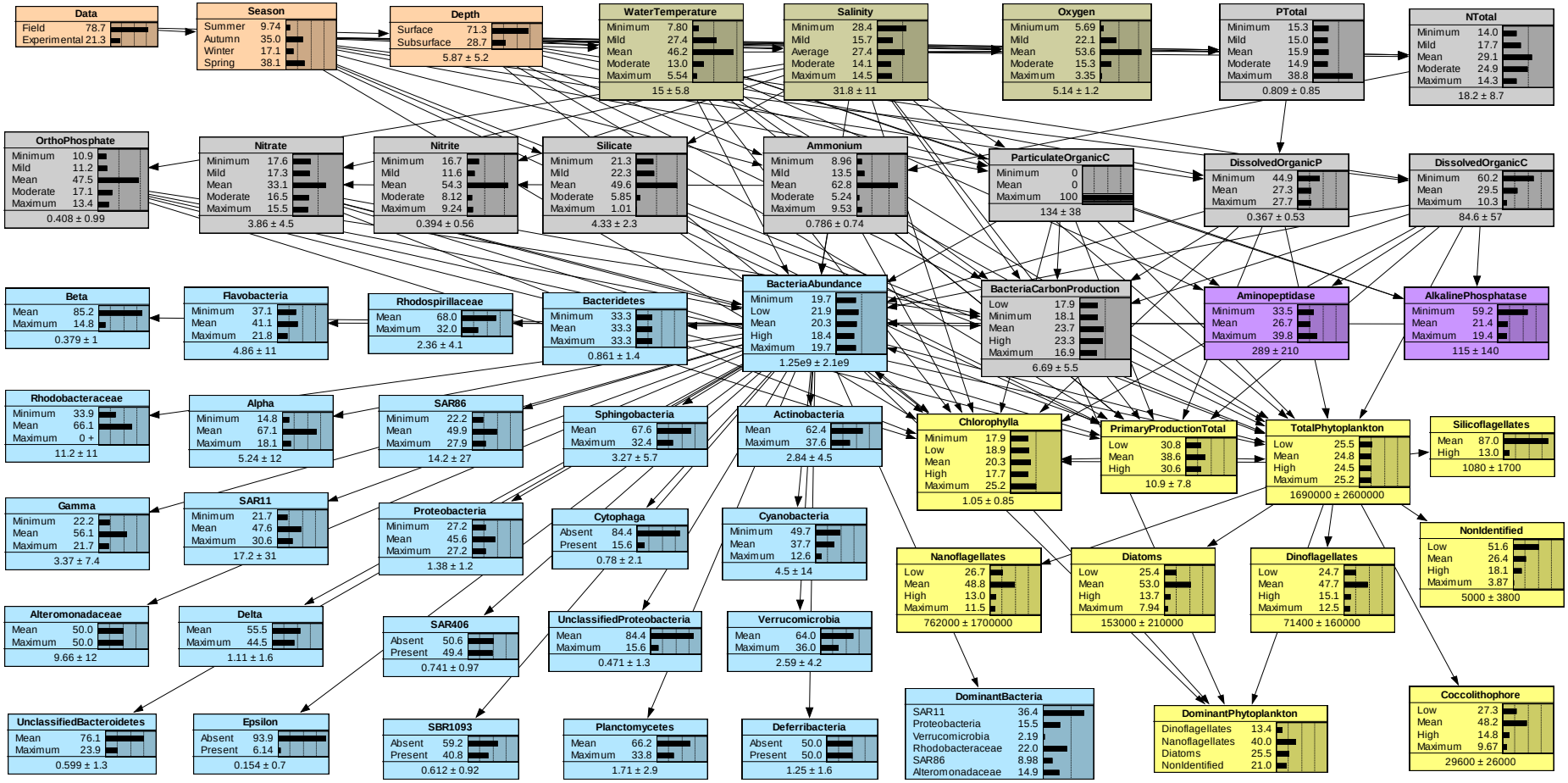
994



995

996 **Fig. S2.3** A scenario test for a nanoflagellate bloom.

997



998

999 **Fig. S2.4** Scenario test for the maximum concentrations of POC in the a priori data to occur.

Supporting Information 3 Sensitivity analysis of the nodes most relevant to the case study presented.

Table S3.1 Sensitivity analysis for POC, dominant phytoplankton and bacterial abundance node**s** (listed to 1% VR).

Output node	**Node**	**VR**
Particulate Organic Carbon (POC)	Aminopeptidase	10.5
	Alkaline Phosphatase	8.3
	Salinity	4.7
	Seawater Temperature	3.3
	Silicate	3.3
	Dissolved Organic Phosphorus (DOP)	3.1
	Total Nitrogen (TN)	2.3
	Total Phosphorus (TP)	1.9
	Dissolved oxygen	1.2
	Nitrite	1.1
Dominant Phytoplankton	Chlorophyll *a*	8.7
	Phytoplankton abundance	3.3
	Coccolithophorids	1.5
	Dinoflagellates	1.4
	Bacteria Abundance	1.2
Bacteria Abundance	Dominant Bacteria	53.2
	Flavobacteria	38.6
	SAR11	37.8
	Deltaproteobacteria	37.5
	Alphaproteobacteria	32.4
	SAR86	27.4
	Actinobacteria	25.8
	Rhodobacteraceae	25.7
	Sphingobacteria	24
	Planctomycetes	22.7
	Verrucomicrobia	22.2
	Cyanobacteria	22
	Gammaproteobacteria	20.5

SAR406	15.8
Unclassified Proteobacteria	10.8
Rhodospirillaceae	10.5
Proteobacteria	10.4
Betaproteobacteria	10.3
Bacteria Carbon Production	1.3
Season	1.2

1004

1005

Table S3.2 The full sensitivity analysis for the POC node with all data (VR %), monitoring data (VR MON %) and experimental data (VR EXP %).

Node	VR (%)	VR MON (%)	VR EXP (%)
POC	100	100	100
Aminopeptidase	10.5	11.3	9.56
Alkaline Phosphatase	8.28	9.17	6.49
Salinity	4.74	2.78	28.9
Season	4.21	5.12	1.33
Water Temperature	3.32	3.81	1.33
Silicate	3.29	3.21	9.27
Dissolved Organic P	3.07	1.5	11.5
N total	2.27	1.62	11.4
Ptotal	1.9	0.632	14.8
Oxygen	1.18	1.22	1.86
Nitrite	1.13	0.419	4.85
Dissolved Organic C	0.998	1.25	0.321
Ammonium	0.969	1.41	9.91
Depth	0.739	0.675	7.69E-06
Nitrate	0.732	0.585	4.27
Ortho Phosphate	0.494	0.29	4.33
Bacteria Abundance	0.363	0.294	1.07
Bacteria Carbon Production	0.285	0.203	1.08
Data	0.203	0	0
SAR11	0.191	0.153	0.488
Planctomycetes	0.184	0.142	0.546
Verrucomicrobia	0.173	0.13	0.542
Actinobacteria	0.164	0.139	0.372
Rhodobacteraceae	0.164	0.13	0.427
Delta	0.151	0.139	0.288
Flavobacteria	0.142	0.113	0.403
Dominant Bacteria	0.137	0.11	0.381
Cyanobacteria	0.123	0.0985	0.364
Alpha	0.0969	0.0758	0.332
SAR86	0.0905	0.0801	0.188
Gamma	0.0741	0.0612	0.18
Primary Production Total	0.0504	0.0273	0.384
Chlorophyll *a*	0.0417	0.041	0.147
Beta	0.0375	0.0293	0.135
Rhodospirillaceae	0.0342	0.029	0.0621
Sphingobacteria	0.0332	0.0242	0.108
Total Phytoplankton	0.0243	0.0346	0.0186
Unclassified Bacteroidete	0.0243	0.0184	0.0683
Dominant Phytoplankton	0.0217	0.0207	0.0714

Cytophaga	0.0195	0.0185	0.0304
Unclassified Proteobacteria	0.0173	0.00981	0.11
Non-Identified	0.013	0.0212	0.00454
Epsilon	0.0103	0.00783	0.0316
Proteobacteria	0.0093	0.00713	0.028
Dinoflagellates	0.00599	0.0101	0.00872
Coccolithophore	0.00591	0.00841	0.00779
SBR1093	0.00535	0.00411	0.0161
Nanoflagellates	0.0044	0.00758	0.00421
Diatoms	0.00202	0.00351	0.00507
Silicoflagellates	0.00198	0.00232	0.000394
SAR406	0.00188	0.000614	0.0256
Deferribacteria	0	0	0
Alteromonadaceae	0	0	0
Bacteridetes	0	0	0

1008

1009

Supporting Information 4 Posterior probabilities depending on data source

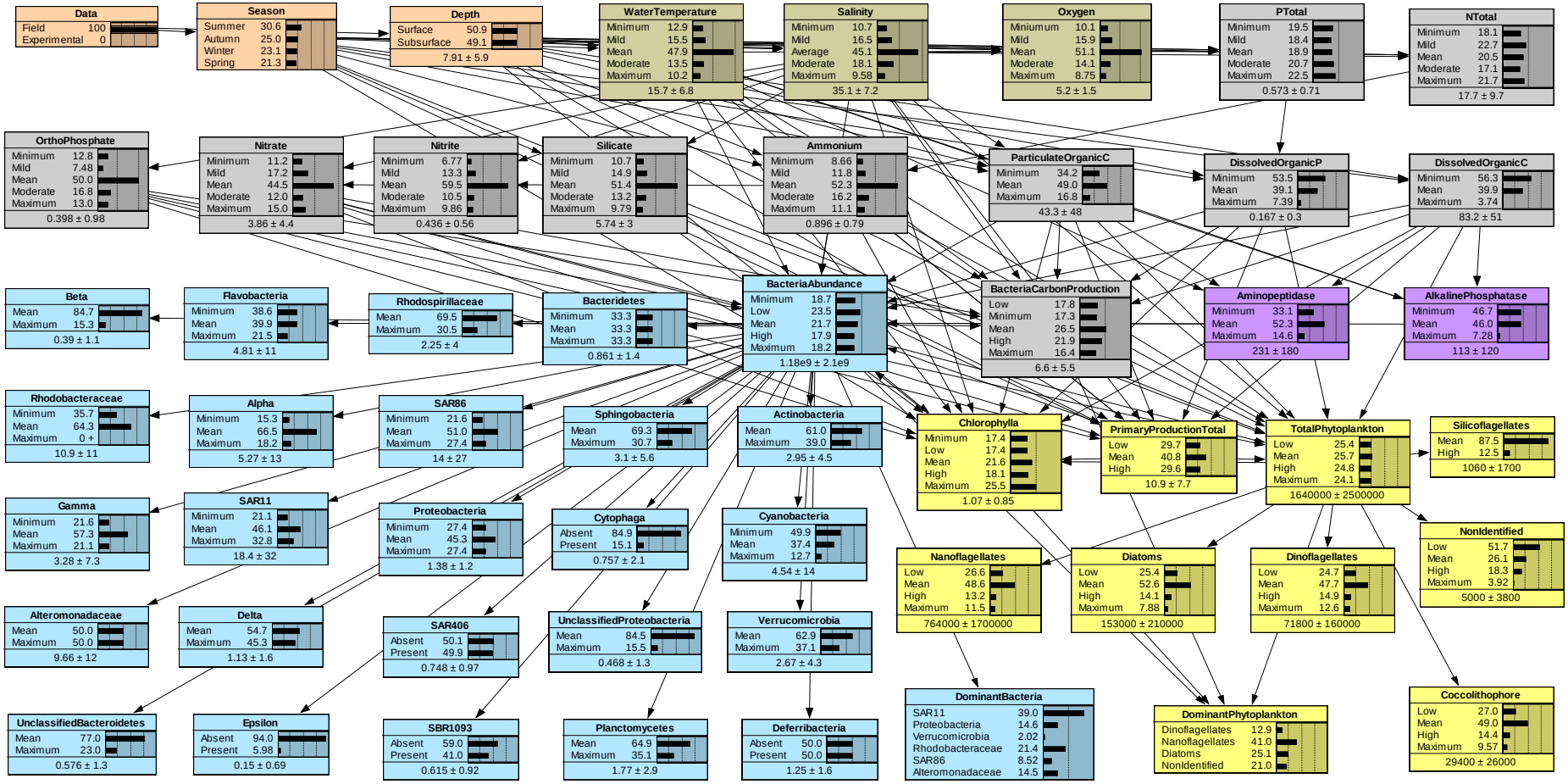


Fig. S4.1 Posterior probabilities informed from field data only.

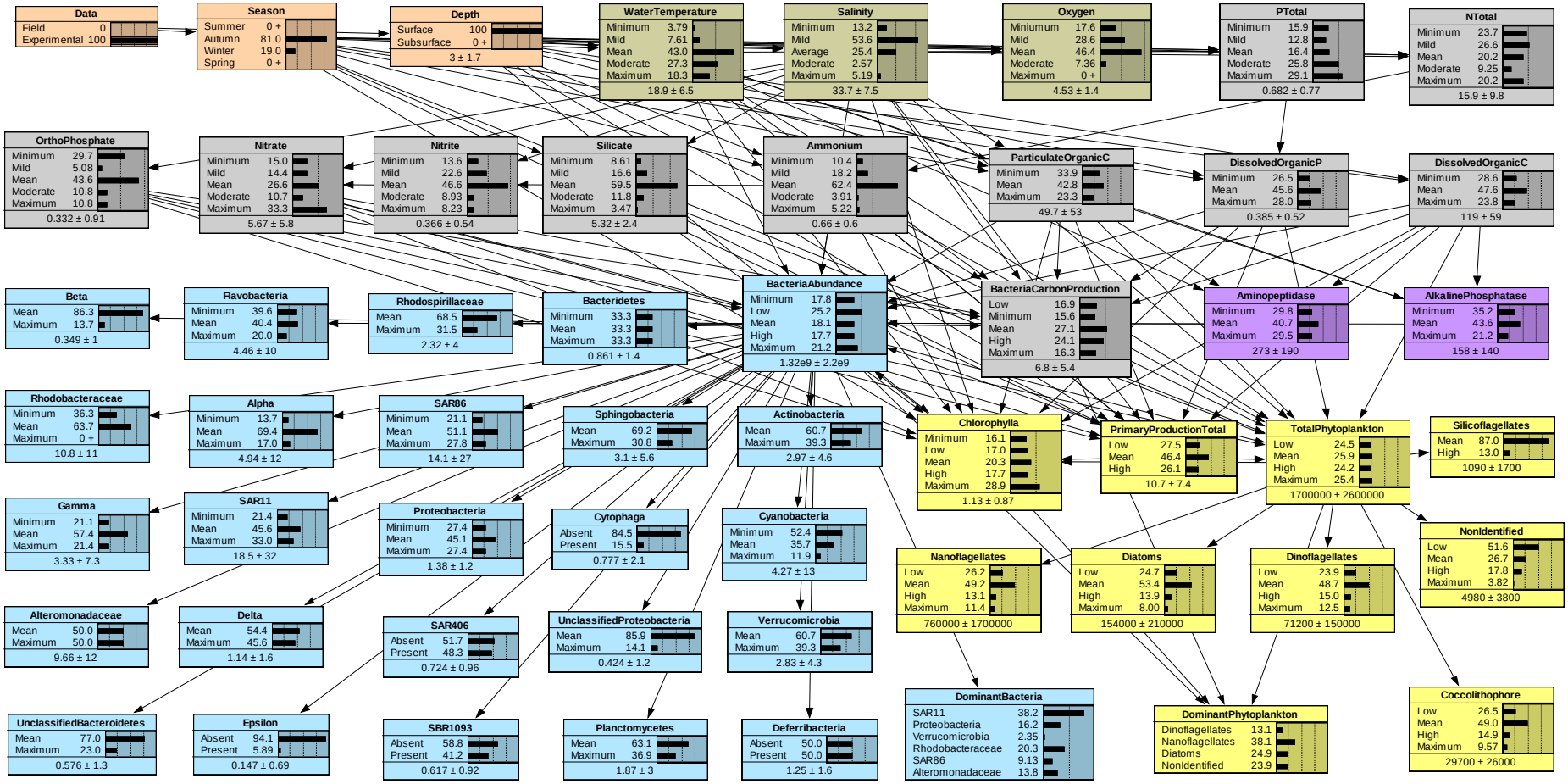


Fig. S4.2 Posterior probabilities informed from experimental data only.