

Hidden Variables in a Dynamic Bayesian Network Identify Ecosystem Level Change

Laura Uusitalo*¹, Maciej T. Tomczak², Bärbel Müller-Karulis², Ivars Putnis³,
Neda Trifonova^{4,5}, Allan Tucker⁴

¹Finnish Environment Institute, Marine Research Centre. Mechelininkatu 34a,
00251 Helsinki, Finland

²Stockholm University, Baltic Sea Centre, 106 91 Stockholm, Sweden

³Institute of Food Safety, Animal Health and Environment BIOR, Daugavgrivas str,
Riga LV-1048, Latvia

⁴Brunel University London, Department of Computer Science, Kingston Lane, Uxbridge,
Middlesex UB8 3PH, United Kingdom

⁵Cooperative Institute for Marine and Atmospheric Studies/Rosenstiel School for Marine and
Atmospheric Science, University of Miami, 4600 Rickenbacker Causeway,
Miami FL 33149, United States

Ecological Informatics

Abstract

Ecosystems are known to change in terms of their structure and functioning over time. Modelling this change is a challenge, however, as data are scarce, and models often assume that the relationships between ecosystem components are invariable over time. Dynamic Bayesian Networks (DBN) with hidden variables have been proposed as a method to overcome this challenge, as the hidden variables can capture the unobserved processes. In this paper, we fit a series of DBNs with different hidden variable structures to a system known to have undergone a major structural change, i.e., the Baltic Sea food web. The exact setup of the hidden variables did not considerably affect the result, and the hidden variables picked up a pattern that agrees with previous research on the system dynamics.

Keywords: Dynamic Bayesian network; hidden variable; ecosystem modelling; Baltic Sea; Gotland Basin

Introduction

Protection of the functioning of ecosystems, as well as safeguarding the ecosystem services delivered to humans, require understanding of the ecological processes and factors that may affect them. Ecosystems are highly complex systems with vast amounts of interactions, feedback loops, and varying time lags due to e.g., different generation lengths of various organisms. This complexity is a major challenge for modelers, particularly as data are often rather scarce due to the relatively high costs of collecting field data, the practical difficulties of collecting samples from all parts of the ecosystem, and the lack of scientific understanding about all factors that may be relevant to the ecosystem functioning.

Ecological modelling has largely focused on statistical models that aim to establish the relationships between a small number of variables, and dynamic, mechanical models that make use of these data-derived (as well as theoretically developed) relationships between

the ecosystem components (Hilborn and Mangel 1997), and more recently, also very large models striving to describe the whole ecosystem (e.g., Fulton 2010, Fulton et al. 2011). However, as such models are parameterized using data, it is usually assumed that the underlying relationships, i.e., the processes that give rise to the data, are unchanging. This assumption might in some cases be untrue, as ecosystems are known to sometimes undergo relatively fast structural changes that have a major effect on the ecosystem dynamics (Scheffer et al. 2001, Möllmann et al. 2008). From the modelling and data analysis perspective, these pose a challenge, since the same functional forms may not describe the relationships between the variables before and after the change (Blenckner et al. 2015).

Identifying such changes is a major challenge, as the natural variation in various observed parameters is often high, making it difficult to separate signal from noise (e.g., Fulton et al. 2003). Further, it is possible that the changes are driven by unobserved variables, i.e., ecosystem components that we do not have data on. Early detection of changes in the ecosystem structure and function would be crucial, however, in ensuring the sustainable use and efficient management of the ecosystems.

Dynamic Bayesian Networks (DBN) with hidden variables have been proposed as a potential method for early detection of structural changes in the ecosystem (Trifonova et al. 2015). Hidden variables in models represent a quantity of interest that cannot be observed directly, or is unobserved for some reason (Murphy 2012). Technically, hidden variables are variables in the model with no data, linked to the observed variables. They can be used to identify e.g. common causes of observed events, such as a disease causing observable symptoms (Murphy 2012).

In a dynamic model, i.e., one that explicitly represents the behavior of the system over time, hidden variables enable the modelling of non-stationary dynamics (Tucker and Liu 2004, Robinson and Hartemink 2009, Ceccon et al. 2011). In other words, they can be

used to detect a change in the interactions of the observed variables over time. When the model parameters are fitted with data, the value of the hidden variable is set so that it maximizes the fit of the model to the data (e.g., the log-likelihood). If the patterns of the observed variables change in the time series, e.g., the slope of a linear correlation between two variables changes, the value of the hidden variable linked to these variables changes. A hidden variable can be linked to multiple, or all, of the observed variables in the model. Then, the hidden variable value depends on all of the observed variables it is linked to, and a change in the pattern of the hidden variable indicates a change in the system interactions. This is potentially highly useful in ecological analyses where complex ecological interactions change in time due to, e.g., global change or changing pressures. However, this possibility is not yet largely explored in ecological analyses, exceptions being Trifonova et al. (2015, 2017), who created spatio-temporal Bayesian networks for the North Sea food web.

In this paper, we evaluate the potential of DBNs with hidden variables in the analysis of a changing ecosystem. The Baltic Sea food web is known to have undergone a major structural change, a regime shift, in the late 1980s–early 1990s (Alheit et al. 2005, Möllmann et al. 2009). The regime shift induced changes in all levels of the food web, particularly in fish and zooplankton (Möllmann et al. 2008). This change in the dynamics of the system is a challenge to ecosystem models whose parameter estimation often relies on the assumption that the interactions don't change dynamically. Our hypothesis is that hidden variables can detect the pattern of the regime shift from the food web data. We explore whether this ability varies between different hidden variable set-ups.

2. Data and methods

2.1 Data

The data (Table 1) cover the higher trophic levels of the food web of the Gotland Basin, Baltic Sea area. They originate from the International Council for the Exploration of

the Sea (ICES), Swedish meteorological and hydrological institute (SMHI), and the Institute of Food Safety, Animal Health and Environment of Latvia (BIOR), and cover years 1975-2012. There was one observation of each variable each year, that is 38 observations in total. The data were log-transformed and standardised to mean 0 and standard deviation 1 in order to achieve normality and enable the model parameter learning.

2.2 Dynamic Bayesian network model

A model of the Gotland Basin food web was built based on expert judgement, and it reflects the current scientific understanding of the key components of the food web and their interactions (Fig. 1). Current methods for learning hidden variables require experts to choose a fixed network structure or a small set of possible structures (Friedman 1997). The experts drafting the model structure were Drs. Bärbel Muller-Karulis, Maciej Tomczak, and Laura Uusitalo, facilitated by Dr. Neda Trifonova (all authors of this paper). The ecosystem experts all have extensive experience from both the Baltic Sea ecosystem and ecological modelling, and the model structure follows the well-known trophic interactions of the Baltic Sea (Arrhenius and Hansson 1993, Österblom et al. 2007, Möllmann et al. 2008, Tomczak et al. 2012, Niiranen et al. 2013). In particular, the model structure includes the predator-prey relationships between cod, herring and sprat (Uzars 1994), the competition between herring, sprat, and juvenile cod for zooplankton prey (Kornilovs et al. 2001, Möllmann et al. 2003, Möllmann et al. 2008), the impact of sprat on juvenile cod via egg predation (Köster and Möllmann 2000), and the dependence of cod recruitment on reproductive volume, i.e., salinity and oxygen conditions in the Central Baltic Sea (Plikshs et al. 1993, Casini et al. 2016). Fishing mortality captures the top-down pressure on the three commercial fish stocks. Furthermore, zooplankton variables are linked to bottom-up changes in productivity and climate via phytoplankton biomass, temperature (Möllmann et al.

2009), and—for *Pseudocalanus*—salinity and oxygen (Otto et al. 2014). The first draft of the model structure was built in an interview between Drs. Trifonova, Müller-Karulis and Tomczak, and modified and finalized through email and skype correspondence between all partners.

Some of the variables in the model are not linked directly to other variables in the same time slice, but are essential in conveying the temporal dynamics between time slices (Fig 2). Five specific hidden variables are included: unobserved juvenile fish age classes of 0, 1, and 3 year old cod, and 0-year-old herring and sprat. Herring and sprat mature and join the spawning stock at the age 2 and cod at the age 4 (ICES 2016); juvenile fish abundances are recorded at the age of 1 in case of herring and sprat and at the age of 2 in the case of food.

The dynamic model is defined in one-year time steps (Fig. 2). The fish-related variables have temporal dependencies across the time steps: the spawning stock sizes (SSB) of the three fish species are autoregressive, as they consist of individual fish that live for multiple years. Each young age class of fish are modelled separately until they reach maturation and join the spawning stock; they exhibit temporal dependency so that the k year old fish are $k+1$ years old in the following time slice, until they reach maturation and join the spawning stock (SSB).

The rest of the variables are assumed not to directly depend on variables in the preceding year. They may, naturally, have temporal autocorrelation due to the fact that variables affecting them are temporally autocorrelated (such as, SSB_{spr} and SSB_{her} exhibit autoregression that may lead to their prey items Ps , Tem , and Ac , to show temporal autocorrelation).

The variables described above form the core of the expert knowledge based model structure. However, it is clear that these variables are not the only ones that affect the

ecosystem dynamics; there are a large number of ecosystem variables and processes which could have an effect on the ecosystem dynamics but which have not been either identified or on which no data exist.

To examine this possibility, three versions of the model were built with different sets of generic hidden variables:

- (1) A model with one generic hidden variable, linked to all other variables in the model in each time slice, and to itself in consecutive time slices (“generic HV”).
- (2) A model with two semi-generic hidden variables: one that is linked to all cod variables (cod fishing, all cod life stages) in each time slice and to itself in consecutive time slices (the “cod HV”), and another similarly linked to all sprat and herring variables and to itself in the consecutive tile slices (the “clupeid 184HV”).
- (3) A model with both the generic HV and the clupeid and cod HVs. The HVs as above; the generic HV is not linked to the cod and clupeid HVs.

Different HV setups were investigated to in order to evaluate whether there is a strong difference between the models depending on how the HVs are linked into the observed variables. We wanted to investigate whether there are patterns behind the cod and clupeid dynamics that can be separated from the general ecosystem change through their specific HVs. Sprat and herring were modelled as depending on the same “clupeid HV”, since the ecology of these two species is rather similar and they are likely to be affected by the same processes. All of the models were parameterized using the same data.

2.3 Model implementation

The models were implemented using MATLAB with Bayes Net Toolbox (BNT, <https://github.com/bayesnet/bnt>). Data analysis was performed and graphs produced in R. The model code written for this work can be found at <https://github.com/luusitalo/BalticFoodWebModel>.

The conditional and marginal probability distributions of the variables were learnt from data using EM algorithm (Lauritzen 1995), using the `learn_params_dbn_em` function of the BNT, which implements EM learning for dynamic BNs (https://github.com/bayesnet/bnt/blob/master/BNT/learning/learn_params_dbn_em.m). All variables were modelled as Gaussian, and the EM learning was initialized by giving the variables priors that were randomly drawn from $N(0,1)$ distribution. The links between the variables were assumed linear, and covariance matrices were assumed diagonal in order to make the parameter learning more robust. This assumption simplifies the model and decreases the number of parameters by assuming that each child's CPD is a linear combination of its parents' CPDs. EM learning is not guaranteed to find the global optimum. Therefore, EM learning was run 100 times for each model, and the model with the highest log-likelihood value was retained for further analyses.

2.4 Inference and analyses

Behaviour of the hidden variables was examined by plotting their time series (mean \pm sd) for all three model versions to see if they show any patterns, and whether those patterns show similarity between the different model specifications. The roles of the hidden variables in the food web were examined by evaluating the weights and signs of the hidden variables on the other variables, to reveal which of the visible variables are most strongly linked to the hidden variables. The weight is a coefficient that is multiplied by the variables in the regressions: For node Y , its continuous valued parents determine the distribution by the

following: $p(Y|X=x \sim N(\mu + Wx, \Sigma))$, where μ and Σ are a given mean and standard deviation for the node, and Wx is the weight given to parent state x .

3. Results

The hidden variables showed distinct patterns in all three model versions (Fig. 3). In the model with only one, generic hidden variable (Fig. 3a), the hidden variable showed a clear decrease between 1975-1981, a distinct increase between years 1981-1990 and a slight decrease towards the end of the time series. In the model with only the fish HVs (clupeid and cod) (Fig 3 b-c), both HVs picked up somewhat similar patterns than the generic HV in the previous model: the increase early in the time series, followed by a stable period with a slight decrease towards the end of the time series (Fig. 3b-c). The clupeid variable featured the dip in the early part of the time series (pre-1980) that was also seen in the generic HV of the previous model.

When all three HVs were included in the same model (Fig. 3d-f), the generic variable (Fig. 3d) showed a rather similar pattern to the model with only the generic HV (Fig. 3a), with the difference that it had a pronounced increase followed by a decrease in the end of the time series. The decrease towards the end of the time series was featured strongly both in the clupeid HV (Fig. 3e) and cod HV (Fig. 3f), in the former, starting from year 2000, the latter in the year 2004, like in the generic variable model (Fig. 3a). (However note that “up” and “down” are completely arbitrary directions in the case of hidden variables; as there are no data, the variable picks up a pattern, but the sign bears no significance. I.e. an increase and a decrease are both simply indicating a change, and their directions can be changed arbitrarily by multiplying with -1. In Fig 3, some of the HV patterns have been inverted this way, to make it easier for human eye to compare the patterns visually.)

In order to evaluate which variables were particularly important to the hidden variable dynamics, the weights of the hidden variables on the other variables of the models

were plotted (Fig. 4). The hidden variables generally had the largest weights on other hidden variables (marked light grey in Fig. 4). Regarding the observed variables, the generic variable had reverse effect on the zooplankton families *Pseudocalanus* (*Ps*) and the other two, ecologically more similar *Temora* (*Tem*) and *Acartia* (*Ac*) (Fig. 4 a-b). The fish-specific hidden variables (Fig. 4 c-f) had the strongest links to fishing mortalities (F) and the young (1-2 year old) year classes, but only a low weight on the spawning stock biomasses (SSB) of the fish species.

4. Discussion

The hidden variables included in the three model versions (Fig. 3) all featured rather similar temporal patterns: most significantly, the steady change in the 1980s, and the decrease in the last three years of the time series. The shift in the HVs in the 1980s (Fig. 3) is concurrent with the regime shift that has taken place in the Baltic Sea food web (Alheit et al. 2005, Österblom et al. 2007, Möllmann et al. 2009). This indicates that the HVs are able to capture the regime shift, particularly if also zooplankton, and not only fish-related variables, are allowed to directly inform the HVs. As a model such as this, once set up, can be run and updated with rather small effort, it could potentially be used to check for possible new changes in the HVs, indicative of major changes in the ecosystem which could then be further investigated.

The generic HV (Figs. 3a, 3d) showed smaller variability than the cod- and clupeid-specific HVs (Figs. 3b-c, 3e-f). In both model setups where the generic HV appears, it is relatively strongly linked with the zooplankton variables *Ps*, *Ac* and *Tem*, while the cod and clupeid HVs were not directly linked to these variables (Fig. 4). These results imply that the zooplankton dynamics, along with the abundance of juvenile cod, are in key role in driving the ecosystem dynamics that are captured by the HVs. This is in agreement with

previous research (Voss et al. 2003, Köster et al. 2005). This, in turn, implies that to detect the HV patterns with low variance, it is important that the HVs are linked to the key observed variables.

When we compare the model with only the generic HV (Fig. 3a) to the model version with all three HVs (Figs. 3d-f), we can see that the generic HV pattern of the latter modes (Fig. 3d) is rather similar to the HV on the former (Fig. 3a), but is missing the decrease in the end. In the latter model however, the cod HV has a steep decrease in the end (Fig. 3f). This decrease may be reflected in the decrease in the generic HV in the model version that does not include the fish-specific HVs (Fig. 3a). This indicates that it might be informative to use few, ecologically relevant HVs rather than just one generic HV.

The cod HV had a similar pattern both in the model with and without the generic HV (Figs. 3c, 3f), whereas the clupeid pattern was somewhat, but less clearly, similar (Figs. 3b, 3e). The clupeid HV had a high variance in all of the model runs, indicating that the processes governing the clupeid dynamics could not be captured by the HV with much accuracy. In the model with only the fish HV (Figs. 3b-c), the clupeid HV shows some similarity to the generic HV as detected in the other model versions. This indicates that the clupeid HV picks up some of the signals that are reflected in the generic HV and which, in the model with all three HVs, are mostly covered by the generic HV.

The variances of the HVs vary largely between HVs and model versions (Fig. 3): the generic HV had a low variance in both model versions (Figs. 3a, 3d), the clupeid HV had a high variance in both models (Figs. 3b, 3e) and the cod HV had a low variance in the model where also the generic variable was present, and higher in the other one (Figs. 3c, 3f). This may partly reflect that the models including the generic HV, which has a direct link to zooplankton, captured the underlying dynamics better than the ones without it.

The fact that the generic HVs had the strongest links with the specific HVs (Fig. 4) follows directly from how the models were fit: since the specific HVs do not have any data, the parameter fitting was free to adjust these variables without any restrictions imposed by the observations. Therefore, these variables tend to vary more in response to the other variables, which generates the stronger link.

The fact that the hidden variables had only a small weight on spawning stock biomasses of all three fish species may be explained by the fact that SSBs are strongly autoregressive, and their response to the environmental factors is more delayed and diluted than that of the other variables. The effect on the SSBs depends strongly on the recruitment of the new year classes to the SSB and on the mortality of the adult fish; the former is mediated through the young year class variables (1-year-old sprat and herring, 2-year-old cod) and the latter largely through the fishing mortality (F), which had stronger links to the hidden variables. The inverse weights on the *Pseudocalanus* versus *Temora* and *Acartia* families were expected based on ecological understanding of these species' environmental preferences (Möllmann et al. 2008).

Marine ecological literature (e.g., Alheit et al. 2005, Möllmann et al. 2009) has identified a regime shift in the Baltic Sea, characterized by two steady states between 1974 and 2005, and a transition in 1988-1993. This change occurs later than the change observed in the current models. Possible reasons for this discrepancy include the fact that the data in this work originated only from the Gotland Basin, where the changes may have started to show earlier than in the whole Baltic Sea data. Secondly, the hidden variable may in fact detect patterns before they can be clearly seen in the observed variables; this may provide a chance for predicting large-scale ecosystem changes. Thirdly, as the hidden variables are modelled as autoregressive in this work, this may restrict how quickly they can change,

meaning that the autoregression forces the beginning of the Change to take place earlier than it would otherwise.

We attempted to find a good model fit by running the EM learning algorithm 100 times and keeping the model with the best log-likelihood; this procedure helps avoid model fits that have ended up stuck in a poor local optimum. However, the parameterizations are not guaranteed to be globally optimal, as the data are quite scarce and the models are rather complex. The temporal patterns of the hidden variables were quite similar on many of the best model fits that were found, however, lending support to the assumption that the models capture the general patterns found in the data.

The models implemented here assumed the original data to be lognormally distributed, although this was not the case with all of the variables; the original data were log-transformed to achieve normality. This is likely to incur a degree of error to the model fitting process. This solution was made as the BNT can only manage continuous variables that are normally distributed, and most of the data fulfilled this assumption. An alternative approach would be introducing mixture models, but this would increase the model complexity, as it would introduce yet more latent variables. In addition, the EM learning of the continuous model parameters assumed linearity between the parents and children, a possibly flawed assumption in the ecological domain, where non-linear “threshold” dynamics are sometimes observed (Muradian 2001, Groffman et al. 2006). On the other hand, Casini et al. (2009) point out that these threshold dynamics may be caused by food web effects, which are directly accounted for in the present models; therefore, the linearity between variables may not be a problem as such.

Another alternative, feasible within the BNT package, would be to discretize the data and build a discrete model. The conditional probability tables would then be learnt from data by the EM learning algorithm without any linearity assumptions. The drawbacks of this

approach, however, are that the break points of the bins are often difficult to determine so that they make ecological sense and are meaningful in the modelling context (Uusitalo 2007). As the data are sparse, the number of bins in each variable must be kept low so that the data will suffice for the estimation of all the parameters. Even so, there will likely be parent-child value combinations that include no or very few data. This leads to instability in the discrete parameter estimation, as a single data point can change the outcome considerably.

Modelling of ecosystem dynamics is challenging due the complexity and, often, relative sparseness, of data. Ecological processes take place in very variable temporal and spatial scales, ranging from very short-term dynamics such as the nutrient uptake by phytoplankton to multi-year processes such as growth and maturation of fish and other long-lived organisms; and from processes varying on spatial scales of meters (such as plankton patchiness, e.g. Horne and Goldman 1994) to processes that occur more or less uniformly on scales of hundreds of kilometers (such as climate). Data are also often available on different scales - such as, fish stock data in one-year resolution (particularly in cold and temperate climates where the reproduction dynamics usually follow seasons) and climate data such as sea surface temperature available even on a daily resolution (based on satellite observations). On the other hand, sometimes observations are scarce even though the process is known to have fast dynamics - an example being plankton data, that may only be collected 1-3 times a year even though these organisms can have strong seasonal cycles and vary considerably on the scale of days to weeks (Horne and Goldman 1994). They, in turn, interact with the ecosystem components with longer life spans, and may for example create bottlenecks that cannot be observed in the data.

This multitude of intertwined scales forces the modeller to make decisions about which dynamics to include and how to model the dynamics that take place within one time step of the chosen temporal resolution. In the current models, the one-year time step

includes temporal plankton dynamics, occurring on variables *Chla*, *Tem*, *Ac*, and *Ps*. In the current expert-opinion based model structure, the decision was made to model these dynamics so that both the primary production (using chlorophyll *a*, *Chla* as its proxy) and the plankton-eating fish (sprat and herring, *SSBSprat* and *SSBHer*) were modelled as the parents of the zooplankton variables (*Tem*, *Ac*, and *Ps*). In reality, these variables affect each other on a shorter-than-year scale, and this model structure, as well as any other taking place on a yearly scale, is a simplification. However, the current model structure was deemed by the experts to be the most reasonable simplification of the system.

It is possible to learn model structures from data, either with or without the presence of hidden variables (e.g. Barber 2012). Structure learning reveals patterns in the data and may be helpful in data mining and prediction, but requires a large amount of data (Barber 2012). As ecological data are often sparse, and on the other hand, there are rather well-developed theories about the relationships of the variables, inclusion of expert knowledge often leads to better models (Alameddine et al. 2011, Julia Flores et al. 2011, Chen and Pollino 2012).

5. Conclusions

The Dynamic BN model is a promising method to analyse complex environmental interactions using a combination of expert and domain knowledge and data bases, and it may help reveal underlying ecological patterns. However, data scarcity can be a problem in model validation. The next stage of this project aims to validate the present model with data from other Baltic Sea basins.

Acknowledgments. This work is an outcome of the BONUS BLUEWEBS project, which has received funding from BONUS (Art 185), funded jointly by the EU, the Academy of Finland, Projektträger Jülich (PtJ), Germany, the State Education Development Agency of

Latvia, the National Centre for Research and Development, Poland, and the Swedish Research Council Formas. MTT was funded by the BalticEye Stockholm University projects. Baltic Eye is partnership between Stockholm University and foundation BalticSea2020.

References

- Alameddine, I., Y. Cha, and K. H. Reckhow. 2011. An evaluation of automated structure learning with Bayesian networks: An application to estuarine chlorophyll dynamics. *Environmental Modelling & Software* **26**:163-172.
- Alheit, J., C. Möllmann, J. Dutz, G. Kornilovs, P. Loewe, V. Mohrholz, and N. Wasmund. 2005. Synchronous ecological regime shifts in the central Baltic and the North Sea in the late 1980s. *ICES Journal of Marine Science: Journal du Conseil* **62**:1205-1215.
- Arrhenius, F. and S. Hansson. 1993. Food consumption of larval, young and adult herring and sprat in the Baltic Sea. *Marine Ecology Progress Series* **96**:125-137.
- Barber, D. 2012. Bayesian reasoning and machine learning. Cambridge University Press.
- Beverton, R. J. and S. J. Holt. 1957. On the dynamics of exploited fish populations. Springer Science & Business Media.
- Blenckner, T., M. Llope, C. Möllmann, R. Voss, M. F. Quaas, M. Casini, M. Lindegren, C. Folke, and N. Chr. Stenseth. 2015. Climate and fishing steer ecosystem regeneration to uncertain economic futures. *Proceedings of the Royal Society B: Biological Sciences* **282**.
- Casini, M., J. Hjelm, J.-C. Molinero, J. Lövgren, M. Cardinale, V. Bartolino, A. Belgrano, and G. Kornilovs. 2009. Trophic cascades promote threshold-like shifts in pelagic marine ecosystems. *Proceedings of the National Academy of Sciences* **106**:197-202.

- Casini, M., F. Käll, M. Hansson, M. Plikshs, T. Baranova, O. Karlsson, K. Lundström, S. Neuenfeldt, A. Gårdmark, and J. Hjelm. 2016. Hypoxic areas, density- dependence and food limitation drive the body condition of a heavily exploited marine fish predator. *Royal Society open science* **3**:160416.
- Ceccon, S., D. Garway-Heath, D. Crabb, and A. Tucker. 2011. The dynamic stage bayesian network: identifying and modelling key stages in a temporal process. *In International Symposium on Intelligent Data Analysis*. Springer, pgs 101-112.
- Chen, S. H. and C. A. Pollino. 2012. Good practice in Bayesian network modelling. *Environmental Modelling & Software* **37**:134-145.
- Friedman, N. 1997. Learning belief networks in the presence of missing values and hidden variables. Pages 125-133 *in ICML*.
- Fulton, E. A. 2010. Approaches to end-to-end ecosystem models. *Journal of Marine Systems* **81**:171-183.
- Fulton, E. A., J. S. Link, I. C. Kaplan, M. Savina-Rolland, P. Johnson, C. Ainsworth, P. Horne, R. Gorton, R. J. Gamble, and A. D. Smith. 2011. Lessons in modelling and management of marine ecosystems: the Atlantis experience. *Fish and Fisheries* **12**:171-188.
- Fulton, E. A., A. D. Smith, and C. R. Johnson. 2003. Effect of complexity on marine ecosystem models. *Marine Ecology Progress Series* **253**:1-16.
- Groffman, P. M., J. S. Baron, T. Blett, A. J. Gold, I. Goodman, L. H. Gunderson, B. M. Levinson, M. A. Palmer, H. W. Paerl, and G. D. Peterson. 2006. Ecological thresholds: the key to successful environmental management or an important concept with no practical application? *Ecosystems* **9**:1-13.
- Hilborn, R. and M. Mangel. 1997. *The Ecological Detective: confronting models with data*. Princeton University Press.

- Horne, A. J. and C. R. Goldman. 1994. *Limnology*. McGraw-Hill, Inc.
- Julia Flores, M., A. E. Nicholson, A. Brunskill, K. B. Korb, and S. Mascaro. 2011. Incorporating expert knowledge when learning Bayesian network structure: a medical case study. *Artif Intell Med* **53**:181-204.
- Kornilovs, G., L. Sidrevics, and J. Dippner. 2001. Fish and zooplankton interaction in the Central Baltic Sea. *ICES Journal of Marine Science* **58**:579-588.
- Köster, F. W. and C. Möllmann. 2000. Trophodynamic control by clupeid predators on recruitment success in Baltic cod? *ICES Journal of Marine Science* **57**:310-323.
- Köster, F. W., C. Möllmann, H.-H. Hinrichsen, K. Wieland, J. Tomkiewicz, G. Kraus, R. Voss, A. Makarchouk, B. R. MacKenzie, and M. A. St. John. 2005. Baltic cod recruitment—the impact of climate variability on key processes. *ICES Journal of Marine Science* **62**:1408-1425.
- Lauritzen, S. L. 1995. The EM algorithm for graphical association models with missing data. *Computational Statistics & Data Analysis* **19**:191-201.
- Möllmann, C., R. Diekmann, B. Müller-Karulis, G. Kornilovs, M. Plikshs, and P. Axe. 2009. Reorganization of a large marine ecosystem due to atmospheric and anthropogenic pressure: a discontinuous regime shift in the Central Baltic Sea. *Global Change Biology* **15**:1377-1393.
- Möllmann, C., G. Kornilovs, M. Fetter, F. Köster, and H. H. Hinrichsen. 2003. The marine copepod, *Pseudocalanus elongatus*, as a mediator between climate variability and fisheries in the Central Baltic Sea. *Fisheries Oceanography* **12**:360-368.
- Möllmann, C., B. Müller-Karulis, G. Kornilovs, and M. A. St John. 2008. Effects of climate and overfishing on zooplankton dynamics and ecosystem structure: regime shifts, trophic cascade, and feedback loops in a simple ecosystem. *ICES Journal of Marine Science* **65**:302-310.

- Muradian, R. 2001. Ecological thresholds: a survey. *Ecological Economics* **38**:7-24.
- Murphy, K. P. 2012. *Machine learning: a probabilistic perspective*. The MIT Press.
- Niiranen, S., J. Yletyinen, M. T. Tomczak, T. Blenckner, O. Hjerne, B. R. MacKenzie, B. Müller-Karulis, T. Neumann, and H. E. M. Meier. 2013. Combined effects of global climate change and regional ecosystem drivers on an exploited marine food web. *Global Change Biology* **19**:3327-3342.
- Österblom, H., S. Hansson, U. Larsson, O. Hjerne, F. Wulff, R. Elmgren, and C. Folke. 2007. Human-induced Trophic Cascades and Ecological Regime Shifts in the Baltic Sea. *Ecosystems* **10**:877-889.
- Otto, S. A., G. Kornilovs, M. Llope, and C. Möllmann. 2014. Interactions among density, climate, and food web effects determine long-term life cycle dynamics of a key copepod. *Marine Ecology Progress Series* **498**:73-84.
- Plikshs, M., M. Kalejs, and G. Grauman. 1993. The influence of environmental conditions and spawning stock size on the year-class strength of the eastern Baltic cod. *Ices Cm* **1**:22.
- Robinson, J. W. and A. J. Hartemink. 2009. Non-stationary dynamic Bayesian networks. Pages 1369-1376 *in* *Advances in neural information processing systems*.
- Scheffer, M., S. Carpenter, J. A. Foley, C. Folke, and B. Walker. 2001. Catastrophic shifts in ecosystems. *Nature* **413**:591-596.
- Tomczak, M. T., S. Niiranen, O. Hjerne, and T. Blenckner. 2012. Ecosystem flow dynamics in the Baltic Proper—Using a multi-trophic dataset as a basis for food–web modelling. *Ecological Modelling* **230**:123-147.

- Trifonova, N., A. Kenny, D. Maxwell, D. Duplisea, J. Fernandes, and A. Tucker. 2015. Spatio-temporal Bayesian network models with latent variables for revealing trophic dynamics and functional networks in fisheries ecology. *Ecological Informatics* **30**:142-158.
- Trifonova, N., D. Maxwell, J. Pinnegar, A. Kenny, and A. Tucker. 2017. Predicting ecosystem responses to changes in fisheries catch, temperature, and primary productivity with a dynamic Bayesian network model. *ICES Journal of Marine Science* **74**:1334-1343.
- Tucker, A. and X. Liu. 2004. A Bayesian network approach to explaining time series with changing structure. *Intelligent Data Analysis* **8**:469-480.
- Uusitalo, L. 2007. Advantages and challenges of Bayesian networks in environmental modelling. *Ecological Modelling* **203**:312-318.
- Uzars, D. 1994. Feeding of cod (*Gadus morhua callarias* L.) in the central Baltic in relation to environmental changes. Pages 612-623 *in* ICES Marine Science Symposia. Copenhagen, Denmark: International Council for the Exploration of the Sea, 1991-
- Voss, R., F. W. Köster, and M. Dickmann. 2003. Comparing the feeding habits of co-occurring sprat (*Sprattus sprattus*) and cod (*Gadus morhua*) larvae in the Bornholm Basin, Baltic Sea. *Fisheries Research* **63**:97-111.

Table 1. Summary of the data			
Variable	Explanation	Source	Unit of measurement
FCod	Fishing-induced mortality of cod	ICES 2013	Instantaneous fishing mortality (Beverton and Holt 1957)
Cod2y	Abundance of juvenile cod, age 2	ICES 2013	Number of individuals, in thousands
SSBCod	Spawning stock biomass of cod	ICES 2013	Metric tonnes
FHer	Fishing-induced mortality of herring	ICES 2013	Instantaneous fishing mortality (Beverton and Holt 1957)
Her1y	Abundance of juvenile herring, age 1	ICES 2013	Number of individuals, in thousands
SSBHer	Spawning stock biomass of herring	ICES 2013	Metric tonnes
FSpr	Fishing-induced mortality of sprat	ICES 2013	Instantaneous fishing mortality (Beverton and Holt 1957)
Spr1y	Abundance of juvenile sprat, age 1	ICES 2013	Number of individuals, millions
SSBSpr	Spawning stock biomass of sprat	ICES 2013	Metric tonnes
Ac	Biomass of zooplankton genus <i>Acartia</i> , measured in the spring	BIOR	mg m ⁻³
Tem	Biomass of zooplankton genus <i>Temora</i> , measured in the spring	BIOR	mg m ⁻³
Ps	Biomass of zooplankton genus <i>Pseudocalanus</i> , measured in the summer	BIOR	mg m ⁻³
Chla	Chlorophyll <i>a</i> concentration in the water, measured in the spring	SMHI, ICES	mg m ⁻³
TSpring	Sea surface temperature, spring	SMHI	°C
TSum	Sea surface temperature, summer	SMHI	°C
RV	Reproductive volume of cod - the water volume in which the cod eggs can survive	BIOR	km ³

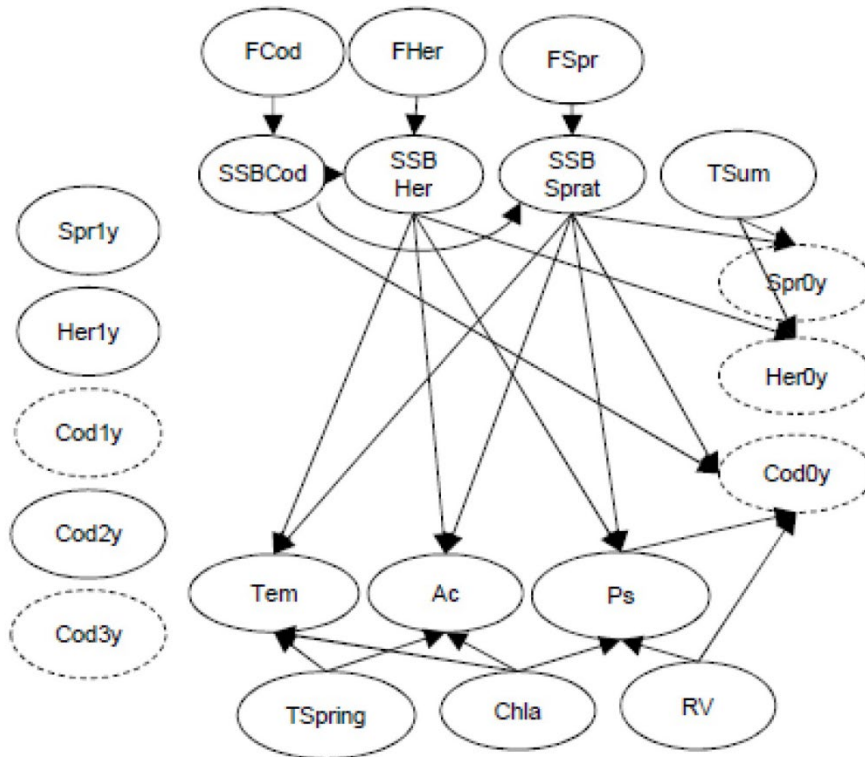


Figure 1. The expert knowledge based model structure, presented without the dynamic aspect. See explanations of the variables in Table 1. The variables shown in dashed line are unobserved, i.e., modelled as the specific hidden variables. Their function is to act as placeholders for the unobserved fish year classes, hence carrying the 2-to-4 year time lags from birth to maturation of the fish. In addition, the model includes 1, 2 or 3 generic hidden variables (depending on the model version), which are not shown.

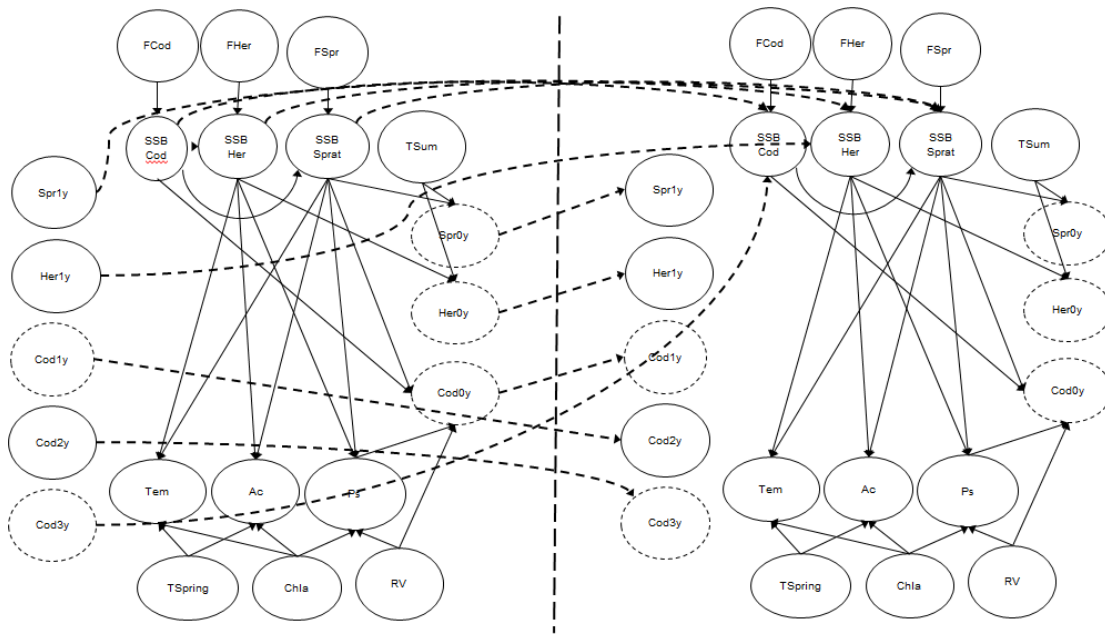


Figure 2. Relationships between the variables across time slices. The vertical dashed line separates the time slices.

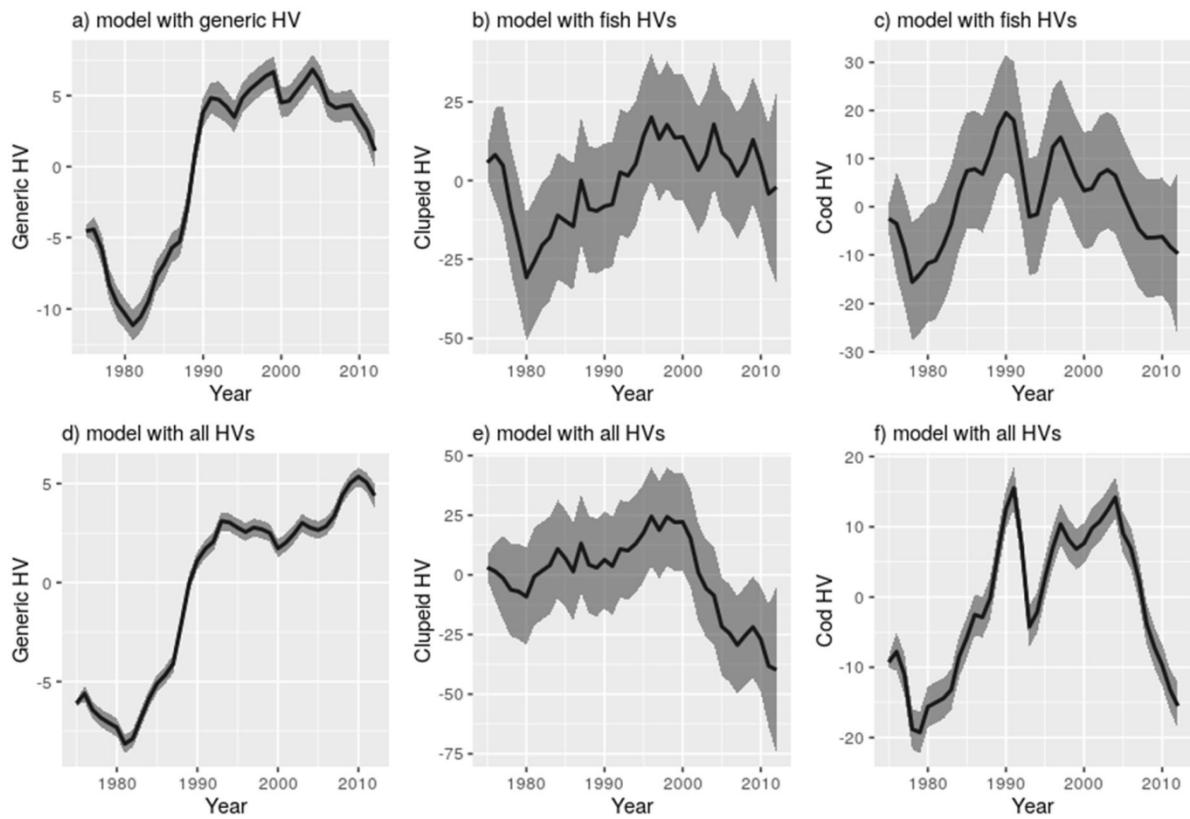


Figure 3. The hidden variable time series. a) the generic HV from the model that includes only this HV; b) clupeid HV from the model with fish-specific HVs; c) cod HV from the model with fish-specific HVs; d) generic HV from the model with all three HVs; e) clupeid HV from the model with all three HVs; f) cod HV from the model with all three HVs.

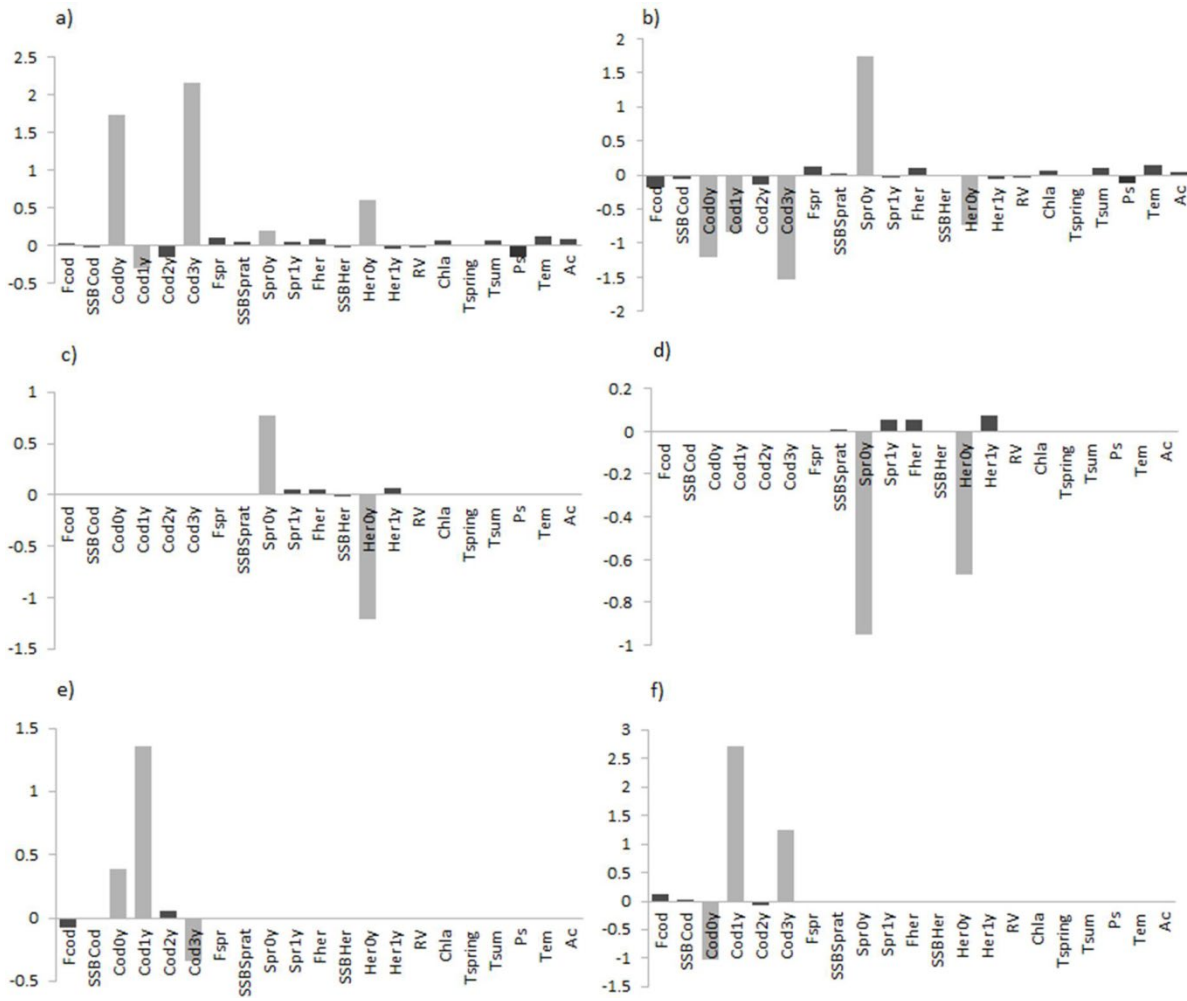


Figure 4. The weights of the generic hidden variables on the other variables in the model. The hidden variables are marked in light grey, the observed variables in dark grey. a) Generic hidden variable, from the model with only the generic hidden variable. b) Generic hidden variable, from the model with all HVs. c) Clupeid HV, from the model with only fish HVs. d) Clupeid HV, from the model with all HVs. e) Cod HV, from the model with only fish HVs. f) Cod HV, from the model with all HVs.