Bayesian Networks to study the factors affecting the population of Pacific Walrus

Adarsh Prakash
Department of Computer Science
University at Buffalo
adarshpr@buffalo.edu | Person #: 5020 8760

Abstract

Population of pacific walrus (Odobenus rosmarus divergens), a species inhabiting arctic and subarctic shelf waters of Chukchi and Bering seas, has been observed to be on a decreasing trend in the last few decades (Fay et al. 1989 and Speckman et al. 2011). This report outlines an attempt to explain the impact of several environmental and anthropogenic factors by utilizing the Bayesian Network constructed by Jay et al. 2011.

1 Introduction

There have been many attempts to study the factors affecting the population of pacific walrus, two of the noteworthy to mention are Fay et al. 1994 and Jay et al. 2011. The former was the first to detail an exhaustive list of environmental factors and anthropogenic stressors. Whereas the latter, was the first to employ Bayesian Networks to model these factors to predict the outcome state of this species through the 21st century. While Jay et al. focused on the final outcome - state of population by estimating various factors and their trends at various points in time, this report focuses on how these factors influence each other and also tries to determine the factors which have larger influence on the population.

Structure of the Bayesian Network from the original paper has been retained. Initially, an attempt has been made to utilize Gibbs Sampling to draw inference. However, after comparing the results with exact inference methods, this sampling technique was found not to be a good estimator of the actual distribution. Hence, all the inference queries have been answered by employing Variable Elimination, a technique for exact inference.

2 Bayesian Network

Figure 1. shows a part of the Bayesian Network detailing the various environmental and anthropogenic factors affecting the population of walruses in Summer/Fall. The nodes colored in blue are the causal variables that influence the output variable. Semantically same factors but with different CPD tables, hence new nodes, form the causal variables for the other two seasons - Winter and Spring. In total, there are 72 different nodes in the network.

2.1 Factors

Each of the environmental factors and anthropogenic stressors as well as their associated nodes are detailed in the following sections. Suffixes *-sumfal*, *-win* and *-spr* in the name of the nodes correspond to *Summer/Fall*, *Winter* and *Spring* seasons respectively.

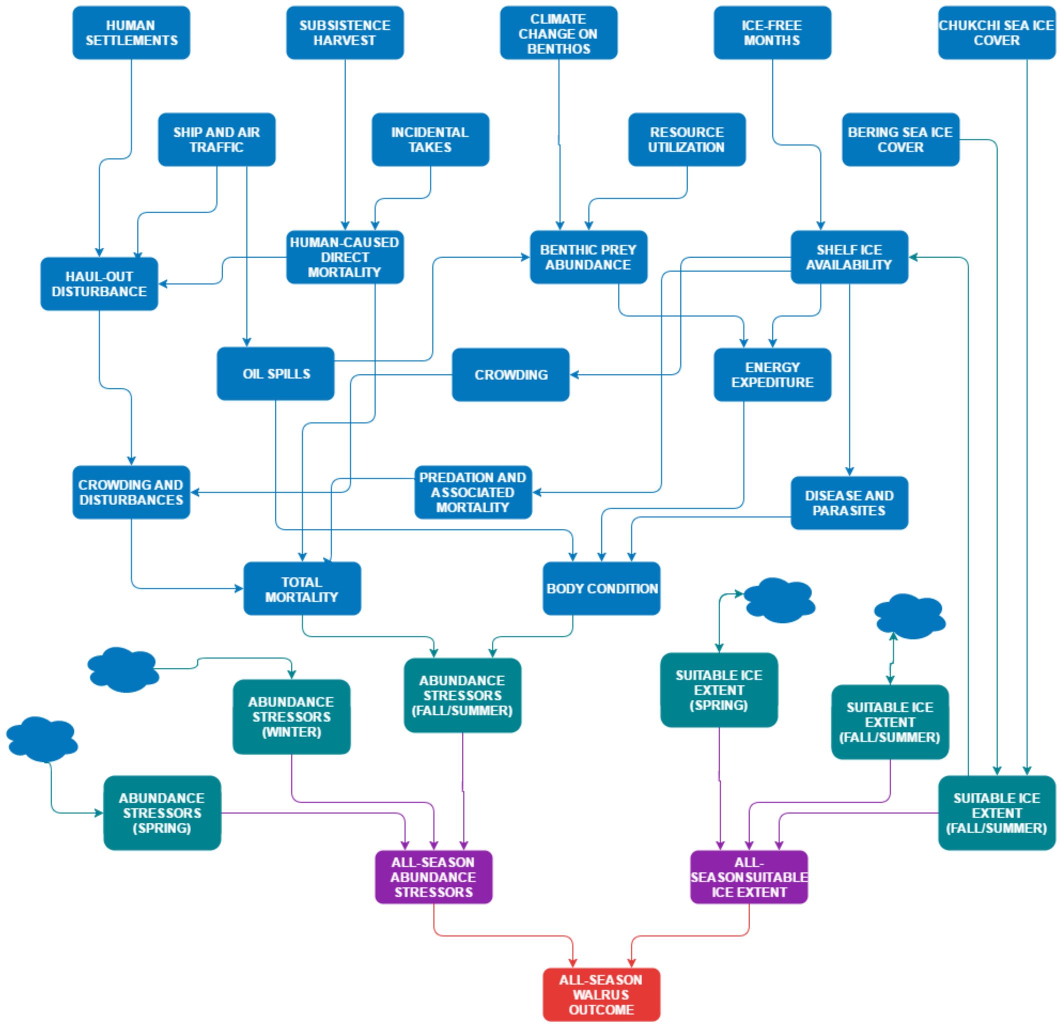

Figure 1: Bayesian Network for Summer/Fall season. The same factors (blue nodes) form the other two-thirds of the network when repeated for Winter and Spring seasons.

2.1.1 Climate change on Benthos *(CCB)*

These nodes represent the cumulative impact of many factors related to climate change effecting the production of benthic prey in all three seasons. A few of these factors include ice thickness, timing of sea ice melt, duration of open water, and ocean acidification.
Nodes: *CCBsumfal*, *CCBwin*, *CCBspr*

2.1.2 Ice free months *(IceM)*

Mean number of months with no sea ice to support hauled out walruses over the continental shelf of the Chukchi and Bering Seas.
Nodes: *IceMsumfal*, *IceMwin*, *IceMspr*

2.1.3 Percentage of Ice Cover *(Ice)*

Extent of sea ice over the Chukchi or Bering Sea in the given season. It is expressed as a percentage of total Chukchi or Bering Sea area over the continental shelf.
Nodes: Bering Sea - *IceBsumfal*, *IceBwin*, *IceBspr*
Chukchi Sea - *IceCsumfal*, *IceCwin*, *IceCspr*

2.1.4 Resource utilization *(RU)*

Impact to benthic prey production from activities that can perturb the seafloor from extraction of natural resources, such as from commercial fishing and oil and gas development. However, it is possible that perturbation to the seafloor at a certain magnitude and frequency could enhance production by releasing nutrients from sediments and by allowing increased recruitment of juvenile organisms.
Nodes: *RUsumfal*, *RUwin*, *RUspr*

2.1.5 Benthic prey abundance *(BU)*

These nodes represent the abundance of benthic prey.
Nodes: *BUsumfal*, *BUwin*, *BUspr*

2.1.6 Energy expenditure *(EE)*

These nodes represent energy expended on foraging and swimming. As benthic prey becomes less abundant or shelf ice is less extensive to provide access to large areas of the continental shelf for foraging, walruses will spend more time swimming to find and forage on prey patches. This might be especially true when walruses (particularly females and young) are forced to use terrestrial haulouts when ice is completely unavailable over the shelf. Walruses might also spend considerable effort swimming in open, rough seas compared to swimming in seas dampened by sea ice.
Nodes: *EEsumfal*, *EEwin*, *EEspr*

2.1.7 Shelf ice availability *(AVA)*

Availability of sea ice to walruses for hauling out during the season. This node reflects two important aspects of sea ice relative to its availability to walruses for hauling out. One is the amount of time no ice is present over the shelf, and hence, the amount of time walruses are forced to use terrestrial haulouts; the second, is the extent of sea ice that is available to walruses. When sea ice is present over the shelf, it can occur in much reduced concentrations and distribution, such as remnant ice in summer, which can affect the distances required for walruses to travel to reach preferred benthic foraging areas.
Nodes: *AVAsumfal*, *AVAwin*, *AVAspr*

2.1.8 Oil spills *(OIL)*

Regularity and severity of hydrocarbons unintentionally released from ship and air traffic, which can affect benthic production by directly fouling benthic organisms or by causing decreased production in the water column and less food fall to the benthos. It can also affect the body condition of walruses through direct contact or indirect bioaccumulation through the food chain.
Nodes: *OILsumfal*, *OILwin*, *OILspr*

2.1.9 Disease and parasites *(DP)*

Incidence of diseases and parasites in the walrus population that would cause a substantial loss of stored energy in an individual.
Nodes: *DPsumfal*, *DPwin*, *DPspr*

2.1.10 Body conditions *(BOD)*

The amount of body reserves an animal possesses, particularly fat and muscle, and can be an indirect measure of reproductive performance.
Nodes: *BODsumfal*, *BODwin*, *BODspr*

2.1.11 Ship and air traffic *(SAT)*

Amount of ship and air traffic from commercial shipping, tourism, commercial fishing, and oil and gas development.
Nodes: *SATsumfal*, *SATwin*, *SATspr*

2.1.12 Crowding and disturbances *(CR/CD)*

Crowding refers to the number of walruses at a haulout. These nodes refer to the intensity of a disturbance on a haulout as functions of the level of walrus crowding and the frequency and magnitude of disturbances on the haulout.
Nodes: *CRsumfal*, *CRwin*, *CRspr*, *CDsumfal*, *CDwin*, *CDspr*

2.1.13 Predation and associated mortality *(PR)*

Number of walruses killed by predators, which are primarily polar bears and killer whales. This does not include human-caused mortalities. It includes walruses killed indirectly from the predator being present and possibly stampeding the herd, which could lead to trampeling.
Nodes: *PRsumfal*, *PRwin*, *PRspr*

2.1.14 Haulout disturbance *(HD)*

Level of disturbances to hauled out walruses on ice and particularly on terrestrial haulouts. Disturbances might increase with levels of ship and air traffic, human settlements near haulouts, and from total human takes.
Nodes: *HDsumfal*, *HDwin*, *HDspr*

2.1.15 Total mortality *(TM)*

Total number of walruses killed by humans, predation, and crowding and disturbance.
Nodes: *TMsumfal*, *TMwin*, *TMspr*

2.1.16 Human-caused direct mortality *(DM/TAO/TAS)*

Total number of walruses directly killed by humans in Russia and Alaska.
Nodes: *DMsumfal*, *DMwin*, *DMspr*, *TASsumfal*, *TASwin*, *TASspr*, *TAOsumfal*, *TAOwin*, *TAOspr*

2.1.17 Suitable ice extent *(SIE)*

This node represents the potential range of walrus movements in the Chukchi and Bering Seas as a function of the % extent of sea ice in the two seas.
Nodes: *SIEsumfal*, *SIEwin*, *SIEspr*, *SIEall*

2.1.18 Abundance stressors *(AS)*

Walrus abundance stressors as a function of body condition and total mortality. Body condition reflects the level of individual fitness and is expected to have an impact on walrus reproduction and survival.
Nodes: *ASsumfal*, *ASwin*, *ASspr*, *ASall*

3 Implementation

The Bayesian Network was modeled in Python using two libraries: **pgmpy** and **libpgm**. This was done to measure and compare performance and accuracy of both these libraries. The following evidential reasoning query, **Q**, was performed on both the libraries for comparison:

Given that Ice Extent ***all through the year*** *was* ***insufficient****,*
What is the probability that Ice Extent in ***Summer/Fall*** *was* ***insufficient ?***

After such a comparison, libpgm was chosen as a better fit for this particular network and further inference tasks are performed using libpgm.

Task	pgmpy	libpgm
Ease of Model Construction	Relatively Hard	Relatively Easy
Time for Exact Inference on **Q**	153.71 seconds	5.64 seconds
Time for Approximate Inference on **Q**	∞ (bug in library)	$\tilde{2}0$ mins for 2500 samples
KL-Divergence of Approximate v/s Exact on **Q**	unknown (because of bug)	0.0084

Table 1: Comparison of both the libraries

4 Choice of algorithm

From Table 1, it is also clear that Gibbs sampling yielded poor results in comparison to exact inference. Table 2 and Table 3 show the deviation of results of Gibbs Sampling from that of Variable Elimination (Exact Inference) technique.

Value	Probability
sufficient in both seas	0.2468
sufficient in one seas	0.3048
insufficient in both seas	0.4484

Table 2: CPD from Gibbs samples (2500)

Value	Probability
sufficient in both seas	0.200448
sufficient in one seas	0.295104
insufficient in both seas	0.504448

Table 3: CPD from Variable Elimination

This deviation between both the distribution obtained from (Gibbs) generated samples and the original distribution can be better expressed through **KL-Divergence**.

$$\textit{KL-Divergence}, KL\ (p||q) = -\frac{1}{N}\ \Sigma_i\{ln\ q(\mathbf{x}_k) - ln\ p(\mathbf{x}_k)\}$$
$$KL\ (p||q) = 0.008382$$

where,

*p(**x**)* is the distribution obtained from the 2500 samples generated using Gibbs sampler

*q(**x**)* is the distribution obtained from exact inference (original distribution)

This comparison was repeated for different sizes of samples. The results are captured in Figure 2. It is evident that the accuracy of Gibbs sampling is still poor and unstable to match the exact inference [1]. Hence, further inference tasks were carried out with Exact Inference, to be specific - Variable Elimination.

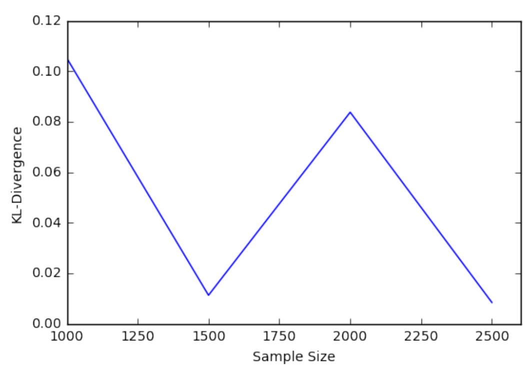

Figure 2: No deterministic trend in KL-Divergence to support inference via sampling

[1]More problems were encountered when tried for higher number of sample sizes (MemoryError). Code from the original library was extracted and many parameters (burn-in, elimination etc.,) were tweaked to force generate more number of samples, but without any fruitful results

5 Inferences

This section poses several queries regarding the factors and stressors affecting the population of walruses. These queries have been answered by the bayesian model using exact inference as described in section 4.

5.1 Seasonal influence

Insufficient ice cover in which season has a stronger influence?

This query tries to evaluate the seasonal fluctuation of ice cover and it's influence on the outcome. Consider a case where low ice cover readings in Summer being passed off as due to normal seasonal change. This query clarifies if any such assumptions hold.

Observed SIEall = "insufficient in both seas"

Query SIEsumfal = "insufficient in both seas" OR
SIEwin = "insufficient in both seas" OR
SIEspr = "insufficient in both seas"

Summary Given insufficient ice cover in both seas all through the year, what are the probabilities that it was because of insufficient ice cover in one of the seasons ?

Results

Season	Probability
Summer/Fall	0.504448
Winter	0.427552
Spring	0.3592

Insufficient ice cover throughout the year is more likely to be caused by insufficient ice cover in Summer/Fall.

5.2 Effect of insufficient ice cover

Insufficient ice cover in which sea has a stronger influence? Chukchi Sea or Bering Sea?

This query tries to isolate the impact of lower ice covers in Chuckchi sea from that of Bering sea.

Observed SIEsumfal = "insufficient in both seas"

Query IceCsumfal = "0 to 10% ice cover" OR IceBsumfal = "0 to 10% ice cover"

Summary Given insufficient ice cover in both seas in a given season, ice cover in which sea is more likely to be the cause ?

Results

Sea	Probability
Chukchi Sea	0.38462
Bering Sea	0.714285

Lower ice cover percentage in Bering Sea is more likely to influence negative trend in population. This also suggests that a larger proportion of walruses inhabits Bering Sea shelf.
(Results consistent across all seasons)

What is the worse that could happen because of insufficient ice cover ?

This query tries to aggregate all the influence that insufficient ice cover can have on the population of walruses.

Observed SIEall = "insufficient in both seas"

Query AllOutCJ

Summary Given insufficient ice cover in both seas all through the year, what is the outcome status of walrus population ?

Results

Outcome	Probability
robust	0.0
persistent	0.150034
vulnerable	0.301984
rare	0.184683
extirpation	0.363299

Compare these results to the effect of anthropogenic factors discussed in further sections.

5.3 Total mortality and indirect anthropogenic factors

What are the indirect impacts of human activities on Total Mortality ?

This query tries to isolate the influence of indirect human factors such as Ship and Air Traffic, Human settlements etc., from that of direct human factors which is determined in subsequent queries.

Observed HUMwin = "high", SATwin = "high"

Query TMwin

Summary Given higher number of Human Settlements and higher levels of Ship and Air Traffic, what is the effect on Total Mortality ?

Results

Mortality Numbers	Probability
low	0.175546
moderate	0.304152
high	0.520302

Compare this distribution to the distribution from next query.
(Results consistent across all seasons)

5.4 Total mortality and direct anthropogenic factors

What are the impacts of direct human threats on Total Mortality?

This query tries to isolate the influence of direct human factors such as hunting, Number of walruses killed from illegal activities and incidentally from fishing, industry, and research activities from that of indirect human factors which is determined in previous query.

Observed TASwin = "veryhigh", TAOwin = "veryhigh"

Query TMwin

Summary Given higher number of legal native hunting (TAS) and higher number of other mortal threats (TAO), what is the effect on Total Mortality ?

Results

Mortality Numbers	Probability
low	0.022830
moderate	0.273480
high	0.703689

Direct human-caused mortal threats have a much greater influence than indirect threats.
(Results consistent across all seasons)

5.5 Isolating the influence of human-caused mortality

Given good environmental conditions - ice extent, if the Walrus Population was extirpated, how likely is it because of high human-caused mortality as opposed to Diseases or other biological factors?

This query tries to isolate the influence of human-caused mortality on population outcome from that of other factors such as Diseases, parasites, body conditions and other biological factors. This is evidential reasoning approach to build on the inference from previous query.

Observed SIEall = "both", AllOutCJ = "extirpation"

Query DMsumfal = "high" OR DPsumfal = "high"

Summary Given sufficient ice cover in both the seas and that the population was still extirpated, how likely is it because of human-caused mortality (DM) or Diseases and parasites (DP) ?

Results

Cause	Probability
high human-caused mortality	0.7572
high mortality due to diseases and biological factors	0.1197

Human-caused mortality is clearly the major threat.
(Results consistent across all seasons)

5.6 Call for action

If Human-Caused Direct Mortality were to be reduced, what is the outcome?

This query tries to evaluate the impact of any actions taken to reduce Human-caused mortality.

Observed DMsumfal = "low", DMwin = "low", DMspr = "low"

Query AllOutCJ

Summary Given reduced numbers of Human-caused Direct Mortality (DM) across all season, what impact on population outcome by the end of that year ?

Results

Outcome	Probability
robust	0.5059
persistent	0.2643
vulnerable	0.1520
rare	0.0479
extirpation	0.0298

Outcome will be in favor of robust and persistent walrus population by a large margin very quickly if effective actions are taken against human-caused mortality!

6 Conclusion

An analysis similar to inferences drawn in Section 5 would be impossible to perform on full probability distribution. To be specific, performing such analysis on this full distribution 72 nodes would require $\mathbf{2}^{72}$ evaluations which is computationally impossible! Probabilistic Graphical Models such as Bayesian Networks can simplify this task to a significant degree.

Additionally, this attempt is a good example to demonstrate the advantages of probabilistic graphical models over conventional machine learning techniques. Although it is possible and elementary to predict the outcome of walrus population using a wide variety of classification methods or other modeling techniques, probabilistic graphical models equip us with the tools to study the relationship between various factors and help make sound decisions about drawn inferences.

Finally, it is clear from the inferences in Section 5 that human-caused direct mortality is a major threat to the population of walruses. It would a great first step to take actions against such activities before focusing on other factors to ensure the sustenance of pacific walrus.

References

[1] Srihari, S.N. (2017) Lecture Slides - CSE 674. NY: University at Buffalo

[2] Koller, D. & Friedman, N. (2009) Probabilistic Graphical Models - *Principles and Techniques.* Cambridge, MA: MIT Press.

[3] Jay, C.V., Marcot, B.G. & Douglas, D.C. (2011) Projected status of the Pacific walrus (Odobenus rosmarus divergens) in the twenty-first century. *Polar Biol (2011)* 34:1065–1084.

[4] Fay, F.H. & Bowlby, C.E. (1994) The harvest of Pacific walrus, 1931–1989. *USFWS R7 MMM Technical Report 94-2.* Anchorage, AK: US Fish and Wildlife Service, Marine Mammals Management

[5] Fay, F.H., Kelly, B.P. & Sease, J.L. (1989) Managing the exploitation of Pacific walruses: a tragedy of delayed response and poor communication. Mar Mamm Sci 5:1–16

[6] Speckman, S.G., Chernook, V.I., Burn, D.M., Udevitz, M.S., Kochnev, A.A., Vasilev. A., Jay, C.V., Lisovsky, A., Fischbach, A.S. & Benter, R.B. (2011) Results and evaluation of a survey to estimate Pacific walrus population size, 2006. Mar Mamm Sci. doi:10.1111/j.1748-7692. 2010.00419.x