

AgriRisk: Bayesian Network modelling for the grape and wine industry of Nova Scotia user guide V0.1

Timothy Lynam

Reflecting Society

March 15th, 2018



Risk Proofing Nova Scotia Agriculture: A Risk Assessment System Pilot (AgriRisk)
Nova Scotia Federation of Agriculture would like to recognize the collaborative relationships that exist among Agriculture and Agri-Food Canada and the Nova Scotia Departments of Agriculture and Environment.

Copyright 2018 ©. No license is granted in this publication and all intellectual property rights, including copyright, are expressly reserved. This publication shall not be copied except for personal, non-commercial use. Any copy must clearly display this copyright.

While the information in this publication is believed to be reliable when created, accuracy and completeness cannot be guaranteed. Use the publication at your own risk. Any representation, warranty, guarantee or conditions of any kind, whether express or implied with respect to the publication is hereby disclaimed. The authors and publishers of this report are not responsible in any manner for any direct, indirect, special or consequential damages or any other damages of any kind howsoever caused arising from the use of this Report.

Acknowledgements

The author gratefully acknowledges the many dimensions of support provided by Nova Scotia Federation of Agriculture (NSFA), Nova Scotia Environment (NSE), Nova Scotia Department of Agriculture (NSDA) and Agriculture and Agri-Food Canada (AAFC). Many people have contributed to the development of the materials that went in to the scenarios workshops and thence the workshops themselves. I am grateful to them all. In particular I am grateful to the many contributions made by Francisco Diez, Rachael Cheverie and Amy Sangster of Perennia, thank you. Joanne Moran (Grape Growers Association of Nova Scotia) and Jerry White (Winery Association of Nova Scotia) have been tremendously helpful and supportive, thank you both. The growers and wine makers of Nova Scotia have been extraordinarily generous with their time and knowledge. Thank you, it makes all the difference in the world. The support provided by Heather MacDougall, Stephen Ford and Ritu Gautam of NSLC has been unprecedented and is greatly appreciated, thank you. Mark MacPherson, Jen Graham and Meredith Flannery have been wonderfully supportive and constructive in guiding the project to fruition, thank you.

Table of Contents

Acknowledgements.....	3
Introduction	6
Using Netica	6
Overview of AgriRisk model suite.....	6
Using the AgriRisk model suite for risk assessments.....	10
Firstly, identify the risk scenario and define the evidence to be used	11
Secondly, create the spreadsheet to capture the results.....	12
Thirdly, implement the scenarios.....	13
Fourthly, use the spreadsheet results to evaluate the results	14
Model updating.....	17
Updating using learning from data	17
Updating using equations	18
Additional resources	19

Table of figures

Figure 1. Overview image of March 2018 version 0.1 BN model of grape growing based on empirical data for price and yield distributions. Blue nodes are associated with the economic components of the model and green the grape growing components. See report noted in footnote 7 for details.	8
Figure 2. Simple winery model with two growers for an examination of risk spreading across the value chain. Grower 1 (top, green nodes) has 100% L’Acadie Blanc on 5 ha. Grower 2 (bottom, yellow) has 60% Chardonnay, 20% L’Acadie Blanc and 20% Pinot Noir. Source: AgriRisk, Winery model v0.1, March 2018.	9
Figure 3. Illustration of the integrated grape and wine model (v0.1) for Nova Scotia. March 2018.	10
Figure 4. Screen shot of paste of links from Netica model to Excel spreadsheet.	13
Figure 5. Screen shot of excel spreadsheet with all evidence from scenarios entered in separate columns.	14
Figure 6. Bar plot of the probability of making a loss under each scenario.....	15

Figure 7. Cumulative probability curve for gross profit / loss for each grower. The x-axis are index values for the qualitative states used in the BN and are shown on the x-axis of Figure 8. 16
Figure 8. Bar chart version of the cumulative probability curve shown in Figure 7..... 17

Tables

Table 1. Experimental design for exploring profitability risk as a function of vineyard size and varietal selection..... 11
Table 2. Bayes factors for comparisons of p(loss) across the four scenario sets..... 15

Introduction

This document presents a brief guide to a) using the Bayesian Network (BN) models developed for the AgriRisk project on risk in the Nova Scotia grape and wine industry; b) updating the models with new information; c) the data and updating the data used to parameterise the models. The BN models were developed using the Netica software package developed by Norsys¹. The assumption in this user guide is that users are familiar with Netica and can navigate and use a BN.

In the next section the reader is directed to tutorials to support their familiarisation with Netica if needed. Experienced users can skip that section. Thereafter a brief overview of the set of models available through AgriRisk is provided. Many of the core elements of these models are similar across the models and therefore, only the core elements will be discussed in relation to model and data updating. Having introduced the models, use of the AgriRisk models is illustrated through case examples where the user can replicate the actions to test that they get the same results. Only the core elements of the models are illustrated in this way.

The model updating section guides the user through basic updating of the models. Only basic updating steps are covered. In a final section a few additional resources that could assist the user are highlighted.

Using Netica

For the purposes of this guide it is assumed that the user has a working knowledge of Netica. If this is not the case it is strongly recommended that the user works through the excellent series of tutorials that Norsys provide. These tutorials are graded from basic to advanced, so the user can select their level of skill and build their knowledge from there.

The Netica tutorials can be found on the Norsys website by following this link:

http://www.norsys.com/tutorials/netica/nt_toc_A.htm

If the link does not work the tutorials can be found (as of March 2018) by navigating the Norsys web site: from the home page, select the “Resources” menu item and then select “Netica Tutorial” from there.

Overview of AgriRisk model suite

The preliminary suite of BN models developed for the AgriRisk project comprises five separate models:

¹ <https://norsys.com/netica.html>

- 1) The base grower model² which is targeted at individual growers;
- 2) The profit maximization decision model based on the base grower model³;
- 3) The loss minimization decision model which is also based on the base grower model⁴;
- 4) The single winery with two growers model which is targeted at small wineries⁵; and the
- 5) Integrated industry level model which is targeted at the whole of industry⁶.

The base grower model is common across all models and is the basic grape growing and grape economics unit or module (Figure 1). At this time the model has two sets of calculations: the set associated with grape growing (green nodes in Figure 1) and the set associated with grape economics (blue nodes in Figure 1). This and other models are described more fully in a companion report⁷.

In the base grower model there are only three important decisions for the user to make⁸ and hence nodes to interact with:

- What varieties to grow and in what proportions;
- What area is the vineyard; and
- What is the year to be used for economic analyses?

All other nodes, whilst the user can select or modify values for them, are calculated by the underlying conditional probability relationships in the model.

Two decision models were developed from the base grower model. The first maximises profit based on selection of a variety. Essentially this model identifies the maximum gross profit (loss) that is likely from a given vineyard size. The only real choice the user has here is vineyard size.

The second decision model, also developed from the base grower model, identifies the maximum of the minimum gross profit (loss) based on selection of a variety. What variety will result in the highest minimum profit or loss? Again, whilst the user can modify the state or distributions of most variables the model is designed to explore the relationships between vineyard size and the probabilities of making a loss given selection of a specific variety.

² NOVA_SCOTIA_GRAPE_GROWER_MODEL_01_EMPIRICAL_PRICE_BRMS_YIELD.neta

³ NOVA_SCOTIA_GRAPE_GROWER_MODEL_01_EMPIRICAL_PRICE_BRMS_YIELD_MAX_PROFIT.neta

⁴ NOVA_SCOTIA_GRAPE_GROWER_MODEL_01_EMPIRICAL_PRICE_BRMS_YIELD_MIN_LOSS.neta

⁵ NOVA_SCOTIA_WINERY_MODEL_01_TWO_GROWERS_ONE_WINERY.neta

⁶ NOVA_SCOTIA_INTEGRATED_GRAPE_WINE_MODEL_01.neta

⁷ Lynam, T. 2018. AgriRisk: Bayesian Network models. Data, analyses, and models. Submitted to NSFA, March 2018.

⁸ Most nodes in the model are modifiable if the user has data or knowledge they seek to enter. With the exception of vineyard size, cost year and variety all nodes have informed prior distributions. These three nodes have uniform (i.e. uninformative) priors.

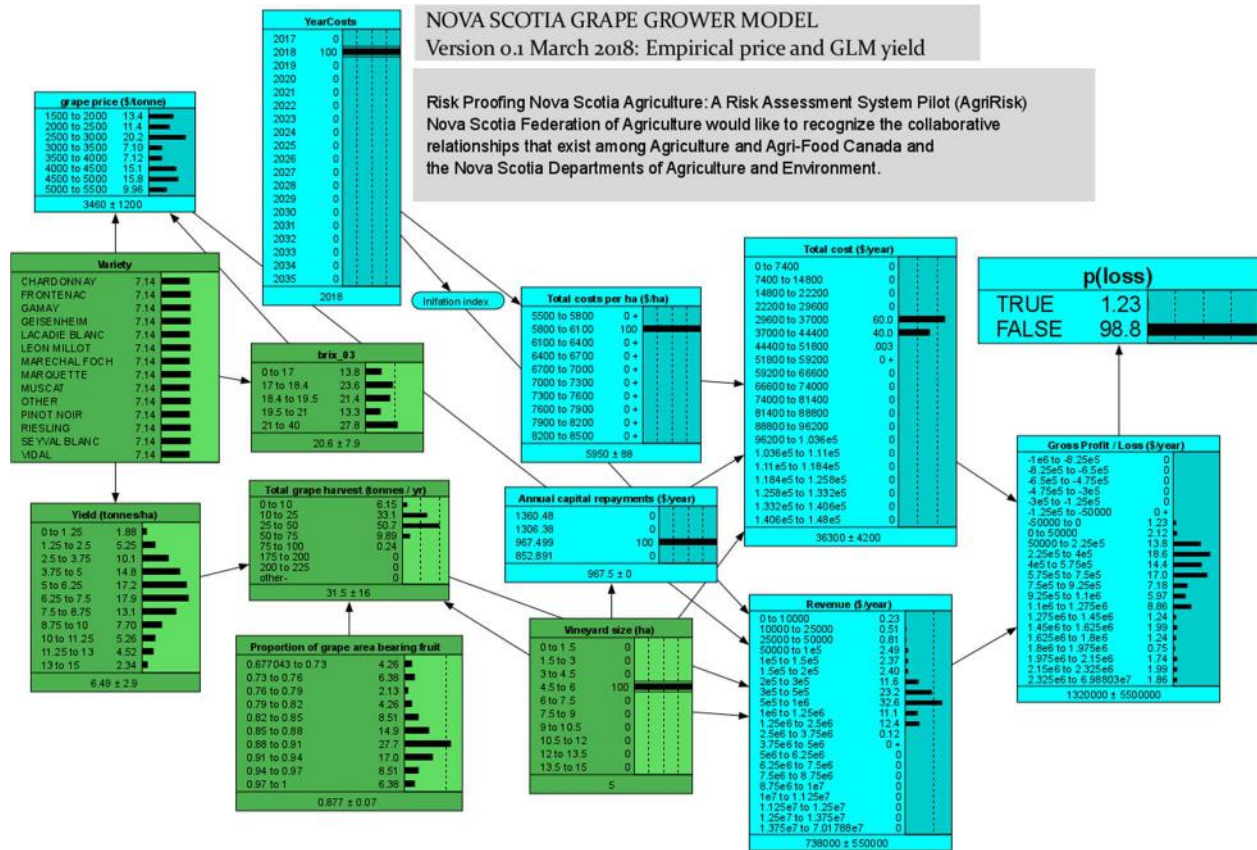


Figure 1. Overview image of March 2018 version 0.1 BN model of grape growing based on empirical data for price and yield distributions. Blue nodes are associated with the economic components of the model and green the grape growing components. See report noted in footnote 7 for details.

The single winery with two growers model (Figure 2) is the most complex of the models and requires the greatest care in its use. The model is constructed with two grower units that are largely independent replicates of the base grower model. The user can choose vineyard sizes and to some extent varieties for these users. Version 0.1 of the model only supports one of the growers having more than a single variety (i.e. one grower must have 100% of their vineyards in a single variety whilst the other vineyard can a combination of varieties). A simple winery is thereafter connected to the two growers. The winery makes single varietal wines from Chardonnay, Pinot Noir and L'Acadie Blanc. The same grape to wine conversion rate is used for each of the varieties. The model estimates the gross profit / loss for each grower and the gross revenue accruing to the winery based on sales of all wine produced to NSLC. This model is on the edge of usability as the conditional probability tables get very large very quickly. The user needs to take care in using it.

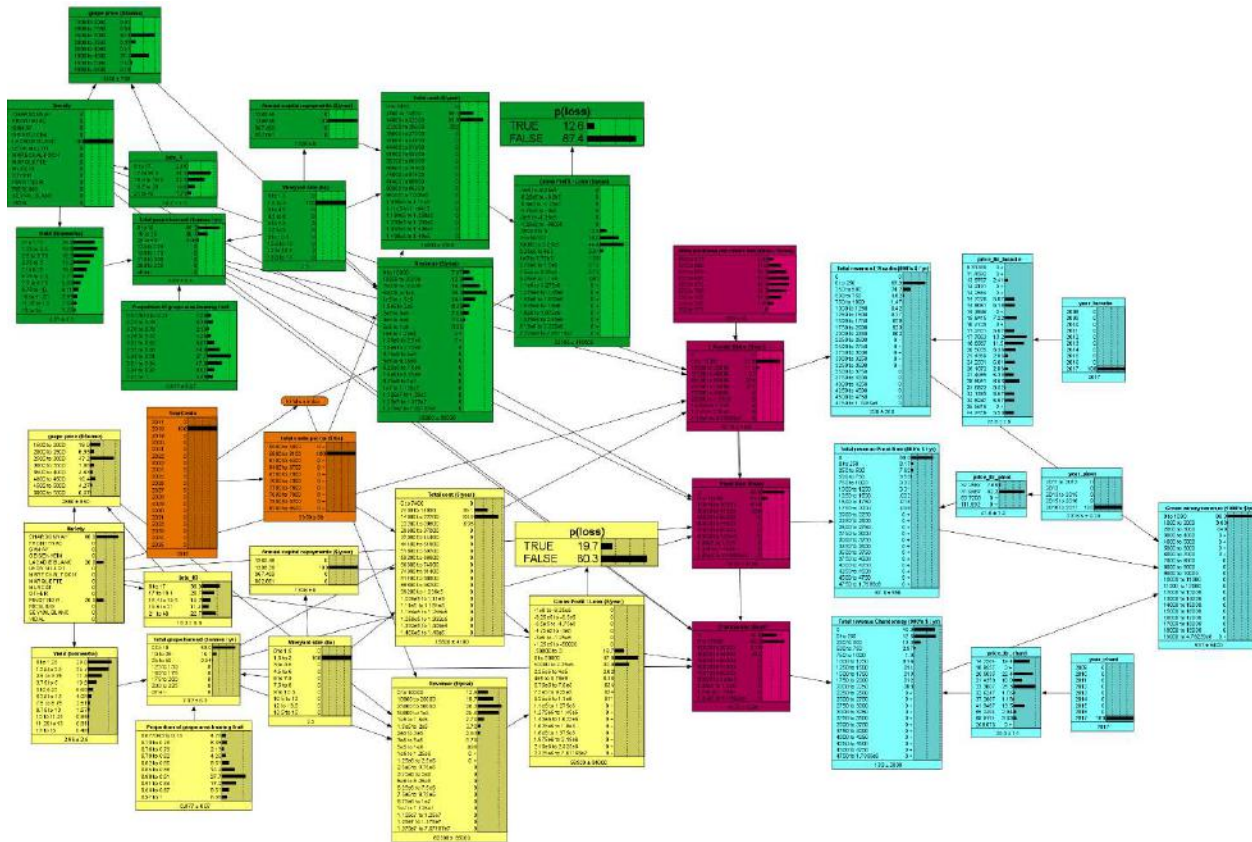


Figure 2. Simple winery model with two growers for an examination of risk spreading across the value chain. Grower 1 (top, green nodes) has 100% L'Acadie Blanc on 5 ha. Grower 2 (bottom, yellow) has 60% Chardonnay, 20% L'Acadie Blanc and 20% Pinot Noir. Source: AgriRisk, Winery model v0.1, March 2018.

The last model in the set is the integrated industry model (Figure 3). In this model the basic grower model is used but adapted for use with larger areas so that the model can estimate the total grape harvest for the whole industry in Nova Scotia. Similarly, the grower economics component of the model aggregates the economic components for all growers in the industry. Thus, instead of a per grower gross profit or loss the integrated model estimates the total gross profit or loss across all growers in the industry. Added to the integrated model is a wine production component (which includes an inventory component) and a wine sales component which is used to estimate the total demand for wine in Nova Scotia. This model can therefore be used to explore supply / demand relationships for wine in general in Nova Scotia. It can also be used to explore the implications of different industry structuring policies.

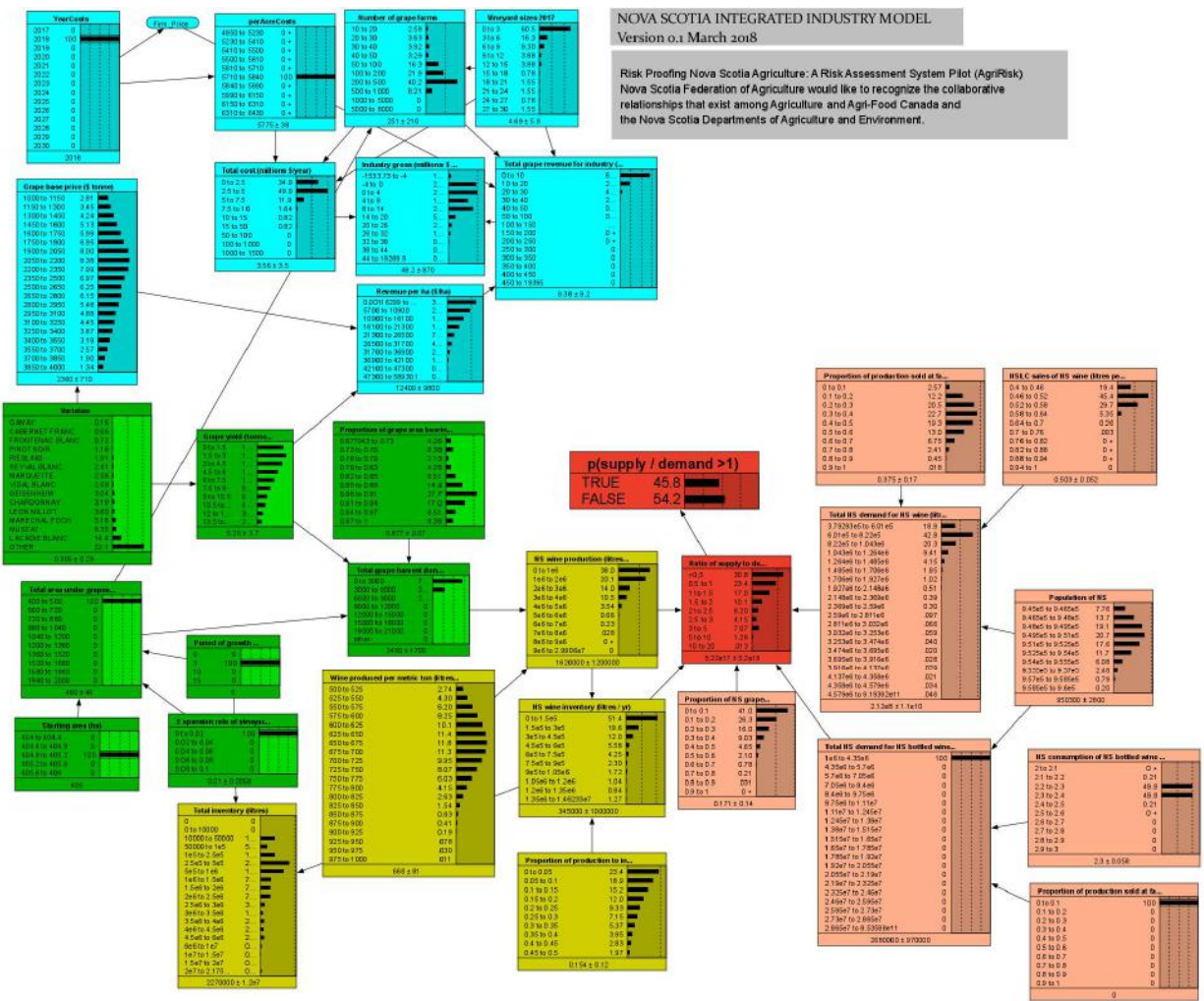


Figure 3. Illustration of the integrated grape and wine model (v0.1) for Nova Scotia. March 2018. Blue nodes are the economic variables; green nodes the grape growing variables; yellow nodes the wine production variables; light orange the sales variables and red the industry supply / demand variables.

Using the AgriRisk model suite for risk assessments

The Netica software makes manipulation of BN models relatively straight forward so the process of using the AgriRisk suite of models is conceptually quite simple:

- 1) Decide on the risk scenario to test and define the settings to be used in the model;
- 2) Create a spreadsheet to capture the results if needed;
- 3) Implement the scenarios by inserting evidence into the model as defined in step 1) above and record the results in the spreadsheet;
- 4) Analyse the results and develop a report or communication of the results.

This process is often extended through iterative analyses or re-analyses as insights from the first analysis prompt questions that can then be answered with subsequent analyses.

By way of an example we will explore the risk of a grower making a loss using the basic grower model.

Firstly, identify the risk scenario and define the evidence to be used

The basic grower model presents us with choices of vineyard size (from 0 to 15ha); varieties (any of the 14 target varieties as well as “Other”) and the year in which we seek to make the evaluation. To simplify matters we select 2018 as the costing year and leave it at that for the rest of the example.

Both variety and vineyard size are presented as uniform distributions which means the base model claims to know nothing about either. This means we can explore both as well as their interaction. Are some combinations of varieties more likely to result in a loss for certain vineyard sizes? All possible combinations of vineyard size and varietal mixes is a very large experimental design that is beyond the scope of this example⁹. We will just explore a couple of combinations and call them scenarios. We will create scenarios that look at small and large vineyard sizes in combination with two single varietal strategies and one mixed varietal strategy. Our experimental (scenario) setup would look like Table 1 below with each cell representing a scenario to be evaluated and ultimately compared.

Table 1. Experimental design for exploring profitability risk as a function of vineyard size and varietal selection.

		Vineyard size	
		Small (1.5ha)	Large (10ha)
Varietal mix	L’Acadie (100%)	Small + L’Acadie	Large + L’Acadie
	L’Acadie (75%) & Pinot Noir (25%)	Small + L’Acadie + Pinot	Large + L’Acadie + Pinot

In this table we have the scenarios defined and so have completed step 1. We will examine two outcomes with each scenario: firstly, the probability of making a loss (i.e. $p(\text{gross profit} / \text{loss} < 1 | \text{scenario})$) and secondly the probability distribution of gross profit / loss for the grower.

⁹ We could use what are called fractional factorial designs to reduce the number of experiments we would have to do with the model, but we will not pursue that here.

Secondly, create the spreadsheet to capture the results

Netica has a very useful function that enables us to record the value of every node in the network whilst the network is running. So, when the network is linked to the spreadsheet every click of the network updates the spreadsheet. To use this facility, open an excel spreadsheet and then go to the open (and compiled) base grower model and select the following nodes in the network (hold the Ctrl key and left click):

Gross Profit / Loss (\$/year)
Total grape harvest (tonnes / yr)
Variety
Vineyard size (ha)
p(loss)

Then go to Netica's report menu and select "Links to Paste in Excel". Then click in a cell of the open spreadsheet and use Ctrl-V to paste the links into the spreadsheet. You should see something like the screen image below (Figure 4) after the paste. If you click on the top right cell you will see the path to the Netica model. The left-hand column of information are the states for each variable in the network¹⁰ and the right-hand column are the probabilities associated with each state. As soon as you now click on any evidence node in the network the changes will be immediately made in the spreadsheet. Because of this live link I usually copy the second column of values (the probabilities for each state) and paste them **as values** in an adjacent column and call this new column baseline¹¹. Make sure there is no evidence entered in the network before you do this initial copy of the baseline.

¹⁰ If you are only interested in some states, you do not need to select all nodes but just the ones you are interested in before the paste.

¹¹ Make sure when you do this that the values are copied as numbers

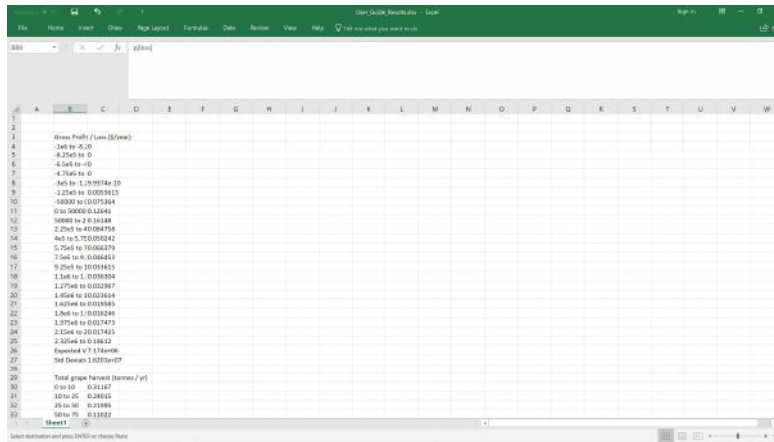


Figure 4. Screen shot of paste of links from Netica model to Excel spreadsheet.

Thirdly, implement the scenarios

Now we are ready to estimate the results for the scenarios. Return to the network and enter the evidence for the first scenario (1.5 ha and all under L'Acadie blanc). Select L'Acadie by just clicking on L'ACADIE_BLANC in the Varieties node of the network. To get the exact vineyard size we are after right click on the Vineyard size node and select Enter Finding->Numeric Value and in the resulting dialogue box enter 1.5. Click OK or hit enter. The evidence is now set so return to the spreadsheet and copy the current column of values to a new column (again be sure to paste as values and numbers) and label the column with the scenario name (I used "LAC_small").

Now return to the network and right click on Vineyard size and then follow the same procedure done above but this time enter 10 in the dialogue box (for 10 hectares). Return to the spreadsheet and copy the current probability results to a new column and name it "LAC_large".

We now need to setup the mixed varietal scenario. Go back to the BN model and right click on the Varieties node and select Enter Finding -> Unknown (Retract) to retract the finding of L'Acadie blanc (alternatively just click on L'Acadie Blanc again and it will remove the finding). Now right button the Varieties node and Enter Finding -> Calibration. A series of dialogue boxes will be presented to you, one for each variety in the node. At each node you will need to enter a value. Enter zero (the number 0) for every state except for L'Acadie Blanc enter 0.75 and for Pinot Noir enter 0.25. The network should now look like the image below. Remember that the BN model is already set up for the large vineyard size we can now go to the spreadsheet and copy the results to a new column and call it something like "LAC_P_large". Now return to the network and set the vineyard size to 1.5 for the "LAC_P_small" scenario and copy the results to a new column. You now have all the results you need to evaluate the scenario. You can if you like close the BN model as we will now work with the results in the spreadsheet. You should have a spreadsheet that looks something like Figure 5 below.

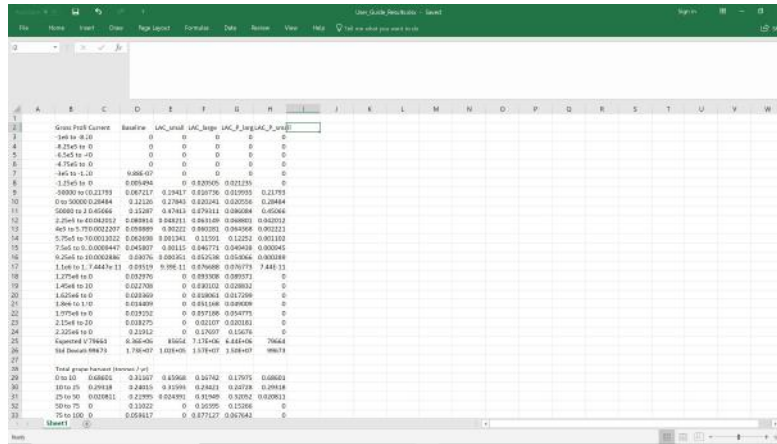


Figure 5. Screen shot of excel spreadsheet with all evidence from scenarios entered in separate columns.

Fourthly, use the spreadsheet results to evaluate the results

First look at the probability of making a loss across the four scenarios. Scroll down to the set of values associated with p(loss). There are a couple of useful ways we can present or discuss these results: One simple way is to plot them as a bar graph: copy the scenario labels down and plot the p(loss) as in Figure 6. It is clear from this graph that small vineyards have a very much greater probability of making a loss. It would also appear (although hard to see clearly) that the LAC_P_large scenario has a slightly higher probability of making a loss.

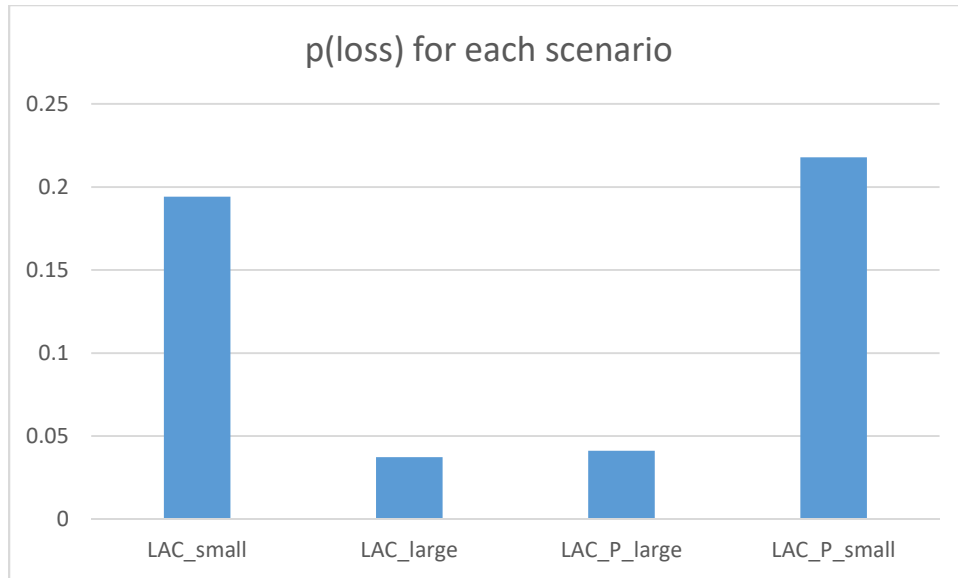


Figure 6. Bar plot of the probability of making a loss under each scenario.

A second way of examining the results is through computing the ratios of probabilities. I think of these ratios as Bayes Factors¹² or the ratios of likelihoods. If the ratio of any two of the probabilities of making a loss is greater than 3 then the difference between those two is worthy of note. The magnitude of the ratio of course depends on which is the numerator and which the denominator. I have created a small table of the key results from the spreadsheet in Table 2 below. From these results it is clear that small vineyards are between 4.7 and almost 9 times more likely to make a loss than are large vineyards (using the two sizes we ran the scenarios for).

Table 2. Bayes factors for comparisons of $p(\text{loss})$ across the four scenario sets.

LAC_small / LAC_large	LAC_small / LAC_P_large	LAC_P_small / LAC_large	LAC_P_small / LAC_P_large
5.21	4.72	5.85	5.29

It is always useful to look at the actual probabilities as well as the Bayes factors. Across all four scenarios the probabilities of making a loss are very small for the large vineyards ($p = 0.03$ or 0.04) but quite considerable for the small vineyards ($p = 0.19$ or 0.22). These probabilities translate to making a loss in about 1 in every 5 years. That is quite a substantial risk!

¹² Kass, R. E., & Raftery, A. E. (1995). Bayes Factors. *Journal of the American Statistical Association*, 90(430), 773-795.

Turning to the probability distributions associated with gross profit or loss. It is often informative to construct cumulative probability curves to examine the likelihood of a discretised continuous variable such as gross profit. Unfortunately excel is a bit difficult to use for cumulative curves as it does not permit qualitative states on the x-axis. So, in Figure 7 and Figure 8 I show two versions of the cumulative curve: the first is a line plot of the cumulative probability and the second a bar chart version. I prefer the line versions as it is easy to see which scenario achieves the greatest gross profit ~ the further to the right a curve is the better the profit. So LAC_P_large and LAC_P_small both have similar gross profit profiles and are better than the other scenarios (although there is a small probability that LAC_small could achieve a higher gross profit). There is much less variability in the LAC_large scenario as it rapidly reaches a maximum gross profit of between \$225,000 and \$400,000. The other scenarios have a slower (i.e. more variable) rise. LAC_small has a higher probability of making a smaller gross profit than LAC_P_small or LAC_P_Large. The combined varietal scenarios generally out performed the single varietal in producing a greater likelihood of higher yields.

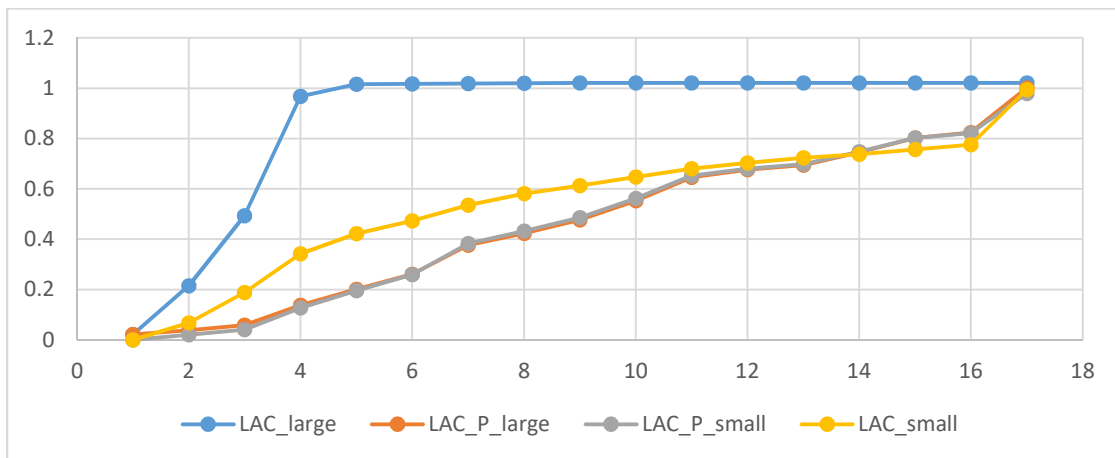


Figure 7. Cumulative probability curve for gross profit / loss for each grower. The x-axis are index values for the qualitative states used in the BN and are shown on the x-axis of Figure 8.

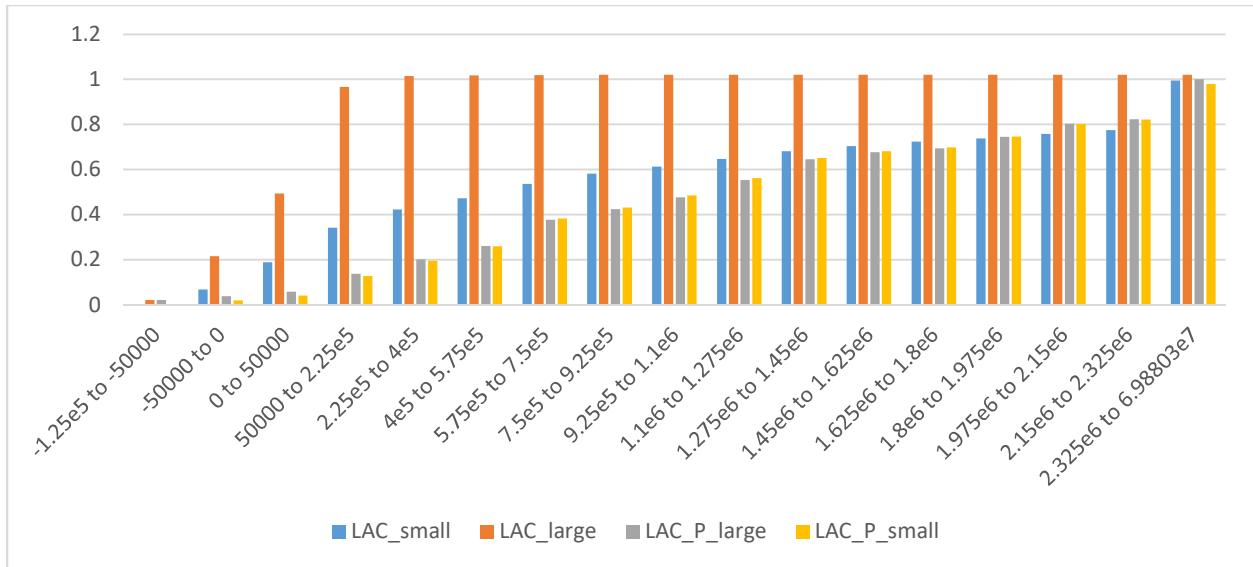


Figure 8. Bar chart version of the cumulative probability curve shown in Figure 7.

That is a simple example of using the base grower BN model to assess risk. It is simple to do and highly informative.

Model updating

There are two ways of thinking about model updating: the first is to do with updating the **structure** of the model and the second has to do with updating the **conditional probabilities** of an existing model. This second type of updating may be called parameter updating. In this guide I will only discuss updating the conditional probabilities. Updating the model structure can be a more complex undertaking and is beyond the scope of this guide.

There are two main types of node relationships used in the AgriRisk BN models: the first are where node relationships (i.e. conditional probability tables) are learnt from data and the second are where equations are used to infer the conditional probability tables. Updating nodes that use the first (learning from data) is relatively straight forward. Updating equations is a little more complex.

Updating using learning from data

As with the scenarios explorations I will use the base grower model to demonstrate updating the model using data. If you double-click on the grape price node and then select the Equation option, you will see that the Equation box is empty. This is because the node was trained using data so the relationships between variety, brix and grape price (i.e. the conditional

probability table) is learnt. If you obtain additional data that can be used to update the variety and price relationships, then updating is fairly simple using the following steps:

- 1) Prepare the additional data in a spreadsheet or text editor so that the data is in a .csv (comma separated values) file. The columns of the spreadsheet should contain the new data. Make sure that the names of each column of data in the spreadsheet that you are going to use is named identically to the name of the variable to be updated.
- 2) Open the base grower model (it is often safest to work with a copy), remove all evidence or findings and select the nodes to be updated;
- 3) Using Cases -> Learn -> Learn using EM¹³ which will result in the software presenting you with a dialogue box to select the .csv file you have just created;
- 4) Follow the prompts, paying close attention to error messages to update the conditional probability tables;
- 5) Recompile the model if necessary.

Netica's messages window will be displayed as the updating occurs and will show you and errors or problems with the updating. It also shows the negative log likelihood and change in negative log likelihood as it updates the model.

Things to pay attention to:

Where you have missing values in the data file Netica expects and Asterix (*). Empty cells can result in odd update values.

Make sure the state names in the data file are exactly the same as the state names in the model. For example, if the data file has a row with LACADIE_BLANC as the variety name but the model uses L_ACADIE_BLANC you will get an error and the updating will not work properly.

Updating using equations

Equations are pervasive across the suite of AgriRisk models. Most of them are fairly simple but some are quite complex with nested if...then relationships. Netica uses two types of equation: probabilistic equations and numeric equations. The base grower model does not have any probabilistic equations. Probabilistic equations are of the form:

$$P(\text{cons_per_cap_NS_wine} |) = \text{NormalDist}(\text{cons_per_cap_NS_wine}, 0.5, 0.05)$$

Which is taken from the integrated model and identifies the probability of cons_per_cap_NS_wine (i.e. NSLC sales of Nova Scotia wine per capita) as being normally distributed with mean of 0.5 and a standard deviation of 0.05. Updating of this form of equation

¹³ The gradient algorithm can also be used but I prefer the EM algorithm.

would usually mean deriving updated estimates of the distributional parameters and entering them in the equation.

Probabilistic equations can be much more complex with complex if then statements but updating them would be a similar process.

Numerical equations can be relatively simple such as the gross profit / loss node in the base grower model which has a simple equation:

$$\text{gross (revenue, total_Cost)} = (\text{revenue} - \text{total_Cost})$$

In contrast the Annual capital costs repayments equation is a little more complex:

$$\text{capitalCosts (vineyard_size)} = (\text{vineyard_size} \leq 2.024)?$$

1360.476 :

$$(\text{vineyard_size} > 2.024 \&\& \text{vineyard_size} \leq 4.05)?$$

1306.383 :

$$(\text{vineyard_size} > 4.05 \&\& \text{vineyard_size} \leq 8.1)?$$

967.499 :

852.891

Updating numerical equations usually means altering the parameters or relationships. For example, in the Annual capital costs repayments equation a user might decide to use different cut-points for the changes in repayment amounts. Or the user may seek to alter the functional form of an equation (division instead of multiplication). But new variables cannot be considered without changing the structure of the model.

Whenever an equation is update the user must convert the equation to a table through selecting the node that has had its equation revised and then from the main menu select Table -> Equation to Table. The user will be presented with a dialogue box asking how many samples per cell. For developing models and testing them I usually accept the default (10) but for operational analyses I would use a much larger number. Larger numbers give more accurate results but are much slower, especially for tables of nodes with many input variables and states.

Additional resources

Netica's help facilities are very informative if a little cumbersome to use. Unfortunately, there does not appear to be a community of users offering support for using Netica as there is

with a lot of other software. I have in the past emailed Norsys for assistance and this is usually forthcoming but not very fast the last time I tried (quite a few years ago). The best bet for additional help are the tutorials and the online help.