Effect of Dataset Size and Auxiliary Data in Bayesian Learning of Advanced Manufacturing: A Composite Autoclave Processing Diagnostic Study

Bryn Crawford, Milad Ramezankhani, Abbas S. Milani*

Materials and Manufacturing Research Institute, University of British Columbia, Kelowna V1V 1V7, Canada

Email: bryn.crawford@ubc.ca, milad.ramezankhani@ubc.ca, abbas.milani@ubc.ca

* Corresponding author

Supplementary Information

1. Example Application of Bayesian Belief Networks for 'Prognostic' Modelling in Composites Manufacturing

The concept of progressively-constrained JPDs acting like a simple look-up table is illustrated in **Table S.1**, where probabilistic-based predictions are used to show the value of using the BBN model, especially in the context of limited information. The CPT of this model has been completed for all 92,160 instances, to help remove the effect of partial or uncertain data present within the CPTs. As the state of the input variables is progressively constrained, the prediction of the output changes. At Step 0, the states are unconstrained and the prediction reflects the distribution of the output, based on the entire dataset. Hence, the output here can be interpreted as stating that for the full dataset, 72% of the 92,160 instances contain exotherm with a class of 0 ("Fail"), whereas 28% of the data contains exotherm with class 1 ("Pass"). At this level, such a statement is very broad and does not provide value beyond a simple counting function. However, once a probabilistic state for the input "part thickness" has been defined, the distribution of the output changes. The ability to define probabilistic states of inputs is also an advantage that BBN models have over many conventional prediction modelling techniques. This is because now, according to the model JPD, by constraining this input, all instances in the dataset that have values for "part thickness" of 0mm, 10mm, 12.5mm, 15mm, 17.5mm and 20mm are all actively excluded as members of the set that we are selecting. Therefore, the distribution of the remaining instances as members of the sub-set will change, which is reflected by a decrease in the likelihood of a class 0 result, and an increase in the likelihood of a class 1 result. This is also a logical conclusion from the perspective of causality in the subject matter domain, as it is known when reducing the thickness of a given composite laminate, the magnitude of the exotherm is also reduced [1].

As another interpretable result, specifically with respect to Step 2, the selection of a tool thickness that is greater than the part thickness is likely to reduce the magnitude of the exotherm and increase the likelihood of a "Pass" outcome. From a physics-based perspective, that is because the heat-sink effect of the tool absorbing excess thermal energy is able to out-compete the heat that is internally generated within the laminate due to cure, as the thermal mass of the tool is likely bigger than that of the part. Similarly, with the selection of the maximum value for heating rate of 5 °C/min, this larger value means that the net thermal flux in or out of the part during process is likely to be greater, as the temperature differential that drives the heat transfer is made greater. This is effective in both heat-up and cool-down regimes, meaning that both the lag temperature and exotherm temperature are lower when subjected to a higher heating rate. However, practically speaking, a higher heating rate may not be an option for manufacturers to leverage in-situ, due to the potential incompatibility of a higher heating rate with the material system used, or due to the autoclave having limited heating power.

Furthermore, given that BBNs are probabilistic in nature, the predictions can be presented based on distribution of variables of interest. To illustrate this, the data used in this study was converted into a form for regression predictions, where discrete values of the outputs exotherm and lag temperature could be considered. Data was discretized into 30 bins for this purpose as shown in **Figure S.2**. Namely, the figure shows the distribution-based predictions for exotherm temperature, where the input variables have been constrained as per Step 3 of **Table S.1**, in addition to the "Tool Material" parent node being constrained to Aluminum 6061, composite AS4/8552, Invar 36 and 1020-Steel, respectively. As can be seen, the predicted exotherm distributions for different tooling materials are distinct from each other. That is, both the average value of the distribution, as well as the spread of the data changes between these cases. This information is shown in contrast to a plot of empirical data to the right of these distributions, where the exotherm as a function of tool thickness is shown, for the four tool material types. The same ranking of the most to least severe exotherm is present in both approaches; Invar 36 provides the lowest (average) exotherm, followed by steel, aluminum and finally, composite tooling offers the highest average exotherm. This is also in agreement with physics-based modelling of this system, where metals provide superior thermal conductivity and thermal diffusivity compared to composites, which is crucial in the tool acting as a heat sink to dissipate the internally-generated heat from the exothermic cure process.

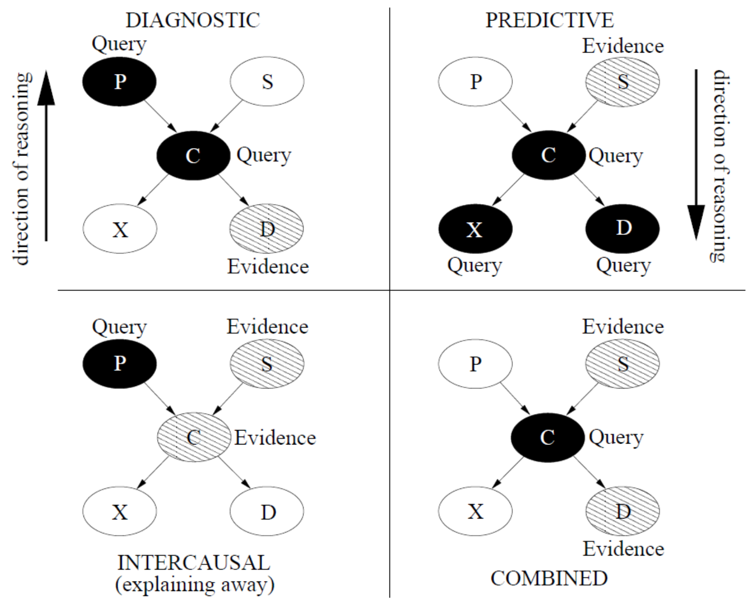
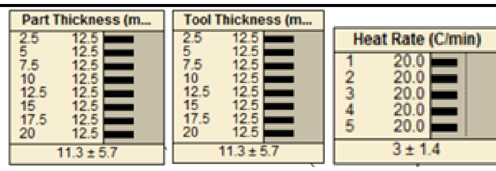
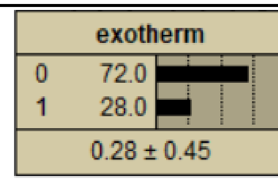
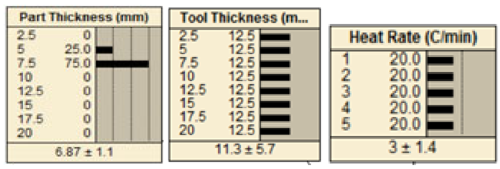
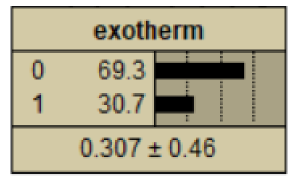
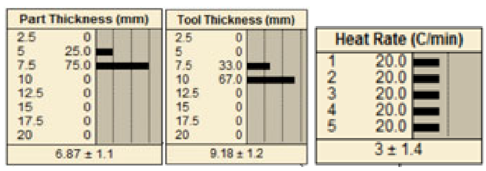
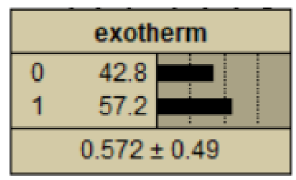
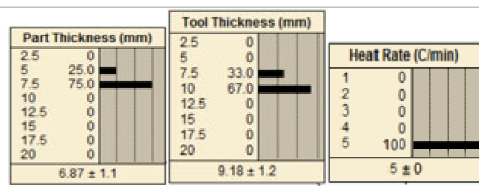
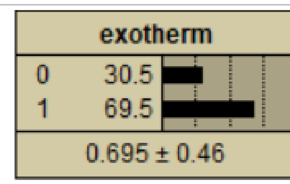


Figure S.1. BBN models have the capability of providing different directions of reasoning; diagnostic, predictive, intercausal and combined queries. Diagnostic queries represent the most straightforward inverse modeling problems, where the state(s) of parent nodes in the shown directed acyclic diagrams are queried, based on both a) the causal relationship between variables encoded by the DAG, and b) the conditional probabilities that quantify the relationships between states [2].

Table S.1. While the JPD of the system is fully defined by the structure of the DAG, it is the values in the node CPTs that define the values of the predicted distributions. As nodes in the network as progressively constrained, the distributions of the remaining unconstrained nodes will become similarly crispier. Here, the predicted class for the output node "exotherm" demonstrates this effect, as the input nodes of "part thickness", "tool thickness" and "heating rate" are sequentially constrained.

Sequential steps to constrain variables	Input State(s)	Prediction of Output State(s)
Step 0: All variables unconstrained		
Step 1: Constrain input "part thickness"		
Step 2: Constrain input "tool thickness"		
Step 3: Constrain input "heating rate"		

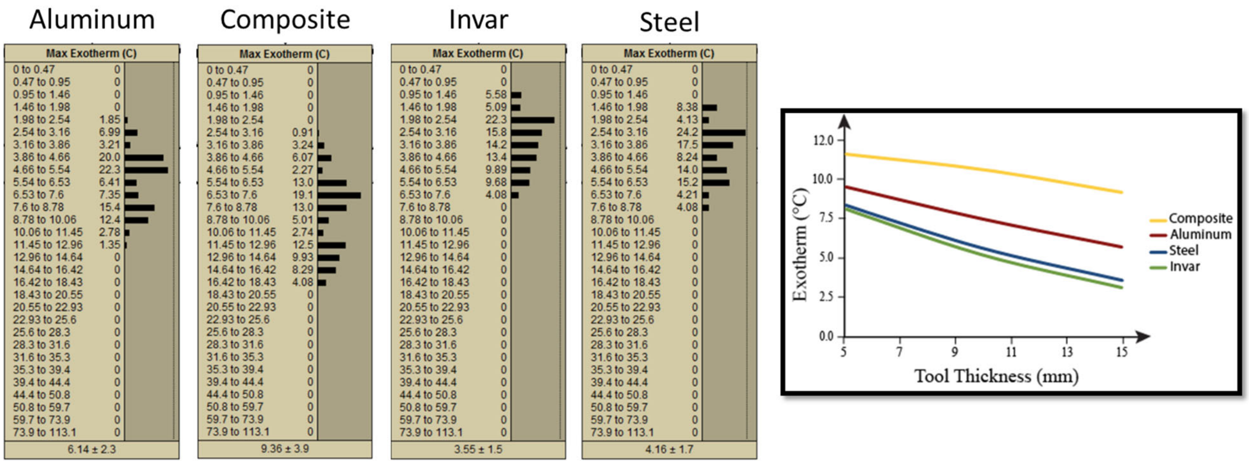
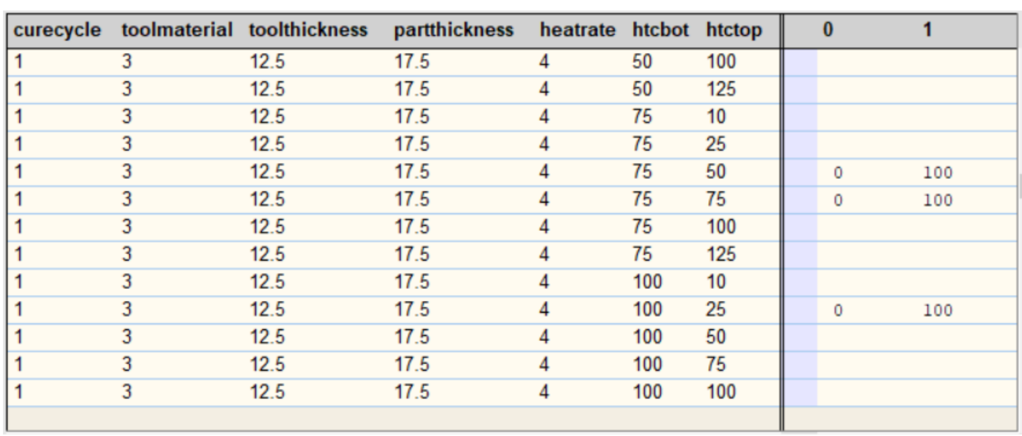


Figure S.2. Illustration of the probability-based predictive capabilities of BBNs. The four distribution-based predictions (left-hand side) of exotherm temperature, where the inputs of part thickness, tool thickness and heating rate have been constrained similar to Table S.1, in addition to the four states of tooling material presented here; aluminum, composite, invar and steel. These predictions are probabilistic in nature and contain more information, when compared to a more deterministic representation of similar data, seen in the plot to the right. The regression-based predictions (right-hand side) provides the same ordinal ranking of the exotherm (e.g. Composites is highest, Invar is lowest), but without any information related to the distribution of the underlying training data.



curecycle	toolmaterial	toolthickness	partthickness	heatrate	htcbot	htctop	0	1
1	3	12.5	17.5	4	50	100		
1	3	12.5	17.5	4	50	125		
1	3	12.5	17.5	4	75	10		
1	3	12.5	17.5	4	75	25		
1	3	12.5	17.5	4	75	50	0	100
1	3	12.5	17.5	4	75	75	0	100
1	3	12.5	17.5	4	75	100		
1	3	12.5	17.5	4	75	125		
1	3	12.5	17.5	4	100	10		
1	3	12.5	17.5	4	100	25	0	100
1	3	12.5	17.5	4	100	50		
1	3	12.5	17.5	4	100	75		
1	3	12.5	17.5	4	100	100		

Figure S.3. The format of the CPT in the Netica software, for entering data about the state of a variable (node) in the BBN, with respect to the states of the associated parent nodes. There are two points of note; i) that there are rows of the table that are empty and if the model is queried about any one of these specific states, the model will be unable to make a prediction, beyond presenting the probabilistic distribution of the output based on the assumption of the model, and ii) that the state of the output "exotherm" is deterministic, as there is no epistemic uncertainty in the numerical model used to generate the dataset.

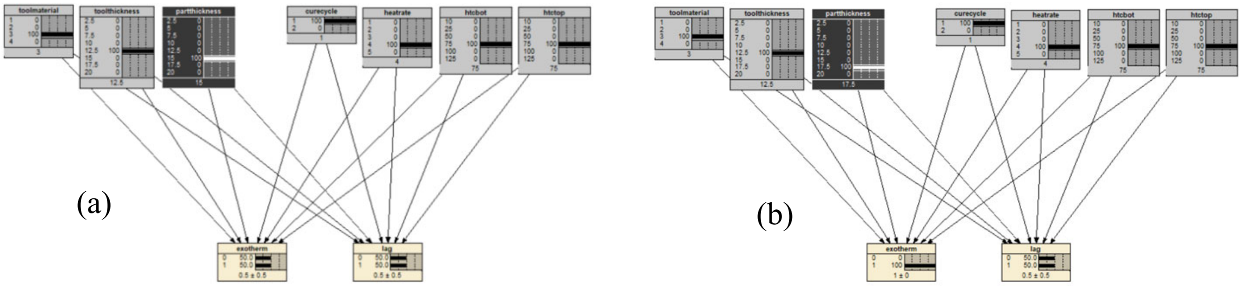


Figure S.4. An example of a major disadvantage for BBNs used as purely predictive tools for specific, deterministic combinations of node states. The model applies the query by fixing the value of all input nodes and the resulting probabilistic distribution of the outputs is given. Here, the (a) left model constrains all inputs and the output is a 50/50 split between both states, which follows the default assumption of a uniform distribution for all variables, in the absence of data. Conversely, the (b) left model applies the same input constraint, except that the state of only one variable (part thickness) is changed, such that this specific state exists within the node CPT. As such, the resulting output is fully deterministic.

Ramp	HTCtop	HTCbot	Exotherm
1.8	70	45	Pass
1.8	70	50	Pass
1.8	70	55	Pass
1.8	75	45	Pass
1.8	75	50	Pass
1.8	75	55	Pass
1.8	80	45	Pass
1.8	80	50	Pass
1.8	80	55	Pass
2	70	45	Pass
2	70	50	Pass
2	70	55	Fail
2	75	45	Pass
2	75	50	Pass
2	75	55	Pass
2	80	45	Pass
2	80	50	Pass
2	80	55	Pass
2.2	70	45	Fail
2.2	70	50	Fail

(a)

Exotherm	Ramp	HTCtop	HTCbot	3.7	3.8	3.9	4	4.1	4.2	4.3	4.7
Pass	1.8	70	45			1					
Pass	1.8	70	50		1						
Pass	1.8	70	55		1						
Pass	1.8	75	45		1						
Pass	1.8	75	50		1						
Pass	1.8	75	55		1						
Pass	1.8	80	45	1							
Pass	1.8	80	50	1							
Pass	1.8	80	55	1							
Pass	2	70	45					1			
Pass	2	70	50					1			
Pass	2	70	55								
Pass	2	75	45				1				
Pass	2	75	50				1				
Pass	2	75	55				1				
Pass	2	80	45			1					
Pass	2	80	50			1					
Pass	2	80	55			1					
Pass	2.2	70	45								

(b)

Figure S.5. The CPTs for (a) the node "Exotherm" and (b) "TCExotherm". In the case of the former, the CPT is composed of 27 entries, based on the three input variables and their states, meaning that the table can be fully completed based on the simulation results. For the latter, there are a total of 54 inputs, as in addition to those for "Exotherm", "Exotherm" itself is an input to "TCExotherm".