Using Bayesian Networks for Structured Learning from Post-Windstorm Building Performance

David B. Roueche[1], Jordan Nakayama[2]

[1] *Auburn University, Auburn, AL, USA, dbr0011@auburn.edu*
[2] *Auburn University, Auburn, AL, USA, jon0003@auburn.edu*

SUMMARY:
Recent advances in post-windstorm reconnaissance have accelerated the amounts of perishable building performance data being collected after extreme windstorms, necessitating better frameworks for knowledge discovery from the data. One particularly promising approach to this need is Bayesian Networks (BN), which have grown in their application in natural hazards research due to their ability to explicitly model causal factors. In this study, a Naïve Bayes Network (NBN) was first developed to observe the influence of wind speed ratio, roof shape, number of stories, roof cover, and pre/post-IBC (2002) on the damage class of a structure and predict the probability of each damage class given a specified scenario. This initial model was derived solely from empirical data and the parameters of influence are modelled with conditional independence, and limiting the model's use. An illustrative hybrid Bayesian Network is also proposed which combines empirical data, known wind engineering theory, and expert opinion to formulate a more holistic model of structural performance in windstorms better suited for parameter inference and building performance predictions.

KEYWORDS: Bayesian Networks, wind engineering, risk assessment, reconnaissance, probabilistic inference

1. INTRODUCTION:

Post-event reconnaissance has long played a role in natural hazards engineering, directly advancing science, policy, and practice. As new data collection techniques and equipment become available, there is an increase in accessible post-event data that researchers and professionals can learn from. An approach that is growing in its implementation in natural hazards research is Bayesian statistics. Bayesian Networks allow for explicitly modeling causal factors by combining the use of prior knowledge and theory, making predictions with incomplete data, and utilizing both subjective and objective data, making them a powerful tool for risk assessment (Fenton and Neil 2018). Specifically, Bayesian Networks have been used to model and learn from wildfires, landslides, and debris flow events (Zheng et al. 2021). The objective of this study is to (1) present a Naïve Bayes Network derived solely from empirical data, (2) propose a Bayesian Network modeling the interdependencies between influence parameters, and (3) postulate the benefits of a Bayesian Network approach, integrating theory and data, for enhancing knowledge discovery from windstorm performance datasets.

2. BAYESIAN NETWORKS:

Bayesian networks (BN) define a joint probability over a set of variables and the corresponding local distributions (Scutari and Denis 2014). BNs are made of two parts, (1) the directed acyclic graph (DAG) and (2) the conditional probability tables (CPT). The DAG depicts the interdependency between variables, or nodes, with arrows connecting the nodes. There are two main types of nodes, the parent nodes, and the child nodes. The child nodes (where the arrow ends) are built from the conditional probability of being in a specific state, given the state of its parent nodes. When a node has no parent, the CPT is the prior probability distribution (Frayer et al. 2014).

Using the DAG and the parameters of influence (**θ**), the joint distribution of all the variables can be factorized into a product of conditional distributions (Scutari and Denis 2014):

$$P(X|DAG,\boldsymbol{\theta}) = \prod_{i=1}^{k} p(X_i|\boldsymbol{X}_{Pa(i)})$$

where X_i is the node which edges are directed to and $\boldsymbol{X}_{Pa(i)}$ is the parent set of nodes.

3. WINDSTORM BUILDING PERFORMANCE DATA

To facilitate learning from windstorm data, the WindStorm Performance Dataset (WiSPD) was used for this analysis (Roueche et al. n.d.). The WiSPD defines the windstorm performance of 4,483 residential structures from four hurricanes and 4 tornadoes. Features of this combined dataset include the record location, building attributes such as year built and number of stories, component-level damage, and design and event-based wind speed estimates for 4,483 records. The illustrative features used in this analysis were the wind speed ratio (ratio of estimated wind speed and design wind speed), roof shape, number of stories, roof cover, and year built (Table 1). The response variable was taken as a damage state, defined as progressive levels of damage (Table 2).

Table 1: Discretized states for the parameters of influence.

Wind Speed Ratio	Roof Shape	Number of Stories	Roof Cover	Year Built
(0.5, 0.7]	Gable	1	Standing Seam	Before IBC (<2002)
(0.7, 0.9]	Hip	1.5	Corrugated	After IBC (<=2002)
(0.9, 1.1]	Combo – Hip & Gable	2	Tile	
(1.1, 1.5]	Complex	3	3-Tab	
(1.5, 2.0]	Other		Laminated	
			Other	

Table 2: Description of damage states for the response node.

Damage State	Description
No damage	0% component damage
Envelope damage	>0% roof or wall cover damage
Structural damage	>0% roof or wall structure damage
Significant structural damage	>25% roof structure damage & >0% wall structure damage

4. CASE STUDY: A NAÏVE BAYES NETWORK

Using the WiSPD, a naïve Bayes network (NBN) was created to observe the effect of the input features on a structure's probability of being in a given damage state. In an NBN the parent nodes are independent of each other as are the marginal probabilities, as shown in the DAG of prior probabilities in Figure 1. An initial sensitivity analysis of the influence parameters showed that the wind speed ratio was the most influential parameter in increasing the probability of exceeding the limit. Using the empirical data and the results of the sensitivity analysis, 5 scenarios (Table 3) were created to observe the changes in the posterior probabilities (Table *4*) for a design level wind event.

Table 3: Description of each scenario used for the predictions.

Scenario	Wind Speed Ratio	Roof Shape	Number of Stories	Roof Cover	Pre/Post IBC
1	0.9 – 1.1	Hip	1	Laminated	Before IBC
2	0.9 – 1.1	Gable	1	Laminated	Before IBC
3	0.9 – 1.1	Hip	2	3-Tab	Before IBC
4	0.9 – 1.1	Gable	2	3-Tab	Before IBC
5	0.9 – 1.1	Combo	1	Laminated	After IBC

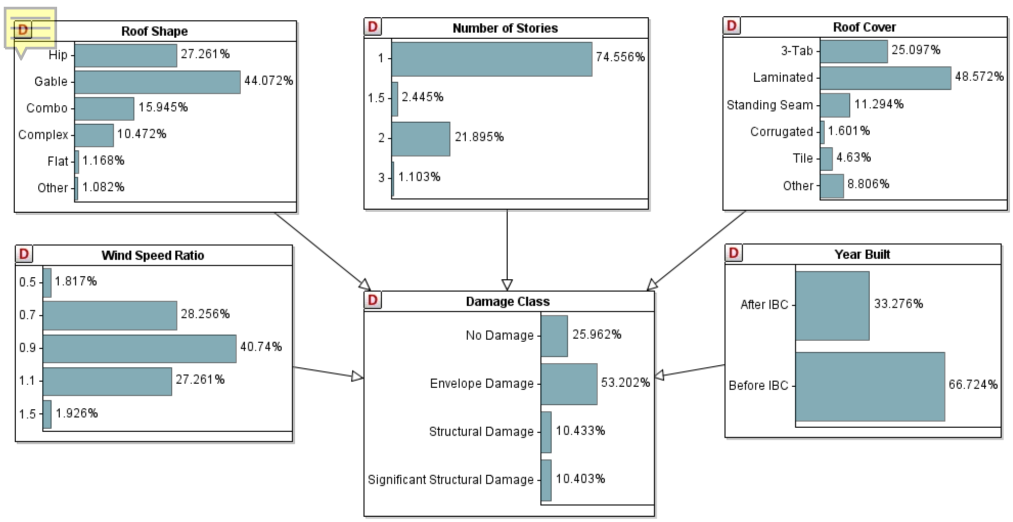

Figure 1: Naive Bayes Network prior probabilities from the WiSPD data.

Table 4: Posterior probabilities for the tested scenarios (ND = No damage, ED = Envelope damage, SD = Structural damage, SSD = Significant structural damage).

Scenario	P(DS = ND)	P(DS = ED)	P(DS = SD)	P(DS = SSD)
1	0.268	0.538	0.100	0.094
2	0.250	0.544	0.088	0.118
3	0.046	0.580	0.206	0.168
4	0.072	0.700	0.078	0.150
5	0.370	0.610	0.012	0.008

5. AN ILLUSTRATIVE BAYESIAN NETWORK

While in the NBN, input features are assumed to be conditionally independent, the actual performance of buildings during a windstorm is driven by a complex interaction between hazard conditions, local site conditions and climatology, building aerodynamics, interactions with adjacent structures, past storms, the local regulatory environment, and even socio-economic factors. These complexities can be better modeled with a Bayesian Network, which allows known relationships and theory, such as the Davenport Wind Loading Chain (Davenport 2011), to be explicitly modeled in the network in tandem with data-driven and judgement-based causal relationships to holistically model wind performance. Figure 2 illustrates a hybrid Bayesian network (HBN) observing a response variable of roof cover damage, but incorporating calculation nodes, continuous nodes, and discrete nodes, and their interdependent relationships.

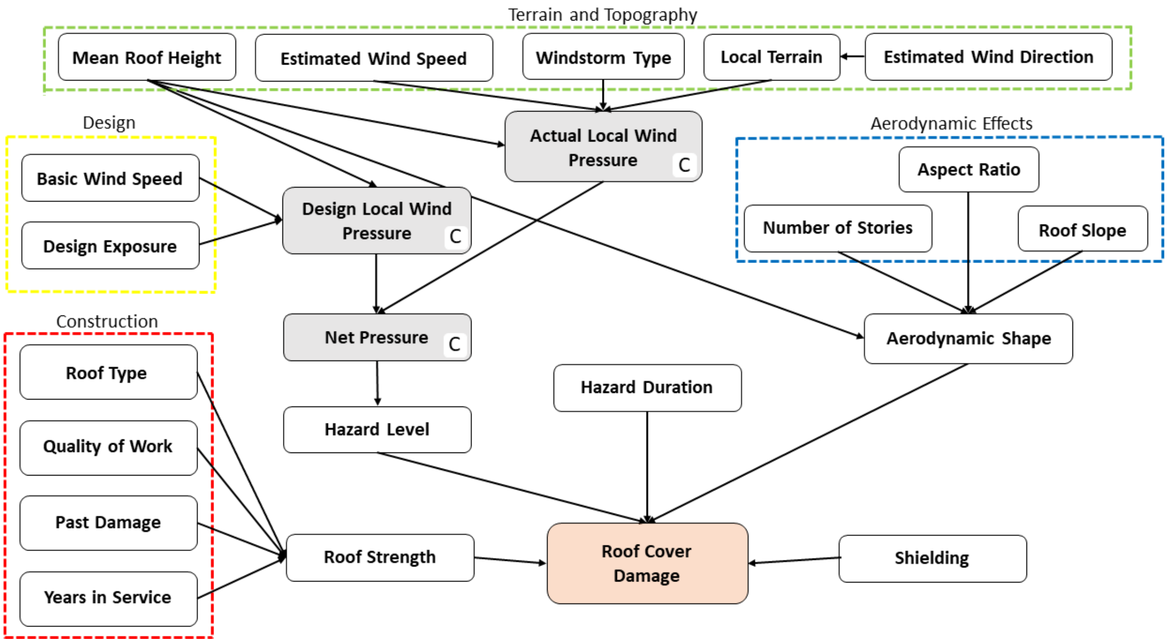

Figure 2: A Bayesian Network with a response variable of roof cover damage where the gray nodes represent calculation nodes.

6. CONCLUSIONS:

The initial naïve network presented in this study was modeled directly from empirical data provided in the WiSPD. However, the parameters chosen for the Naïve Bayes network are not the only parameters that influence the damage class. By incorporating theory and the principles established in the Davenport Wind Loading Chain, the network becomes more complex and requires more information than that provided in the WiSPD. However, using a Bayesian approach to understand structural performance in windstorms allows for a combination of empirical data, theory, and expert opinion to be used to formulate a more holistic understanding.

ACKNOWLEDGEMENTS
This work was sponsored in part by the National Science Foundation under grant CAREER-1944149. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

REFERENCES:

Davenport, P. A. G. 2011. “Announcement of the Alan G. Davenport Wind Loading Chain.” 3.
Fenton, N., and M. Neil. 2018. *Risk Assessment and Decision Analysis with Bayesian Networks*. Chapman and Hall/CRC.
Frayer, J., Z. Sun, D. Müller, D. K. Munroe, and J. Xu. 2014. “Analyzing the drivers of tree planting in Yunnan, China, with Bayesian networks.” *Land Use Policy*, 36: 248–258. https://doi.org/10.1016/j.landusepol.2013.08.005.
Roueche, D., J. Nakayama, T. Kijewski-Correa, and D. Prevatt. n.d. “A Unified Multi-Event Windstorm Performance Testbed for Single-Family Residential Buildings.” *Nat. Hazards Rev.*, ((under review)).
Scutari, M., and J.-B. Denis. 2014. *Bayesian Networks: With Examples in R*. CRC Press.
Zheng, Y., Y. Xie, and X. Long. 2021. “A comprehensive review of Bayesian statistics in natural hazards engineering.” *Nat. Hazards*, 108 (1): 63–91. https://doi.org/10.1007/s11069-021-04729-2.