Wildfire Prediction: Handling Uncertainties Using Integrated Bayesian Networks and Fuzzy Logic

Mohsen Naderpour[1,2], Hossein Mojaddadi Rizeei[2], Fahimeh Ramezani[1]

[1] Centre for Artificial intelligence (CAI)
[2] Centre for Advanced Modelling and Geospatial Information Systems (CAMGIS)
Faculty of Engineering and IT, University of Technology Sydney (UTS), Ultimo, NSW 2007, Australia
{Mohsen.Naderpour, Hossein.MojaddadiRizeei, Fahimeh.Ramezani}@uts.edu.au

Abstract— **Wildfire is one of the most frequent natural hazards across the globe, one which has cast a malevolent shroud over many communities in recent years, causing significant risk to human lives, infrastructure, and property. Wildfires are hydrogeological events which are bound to escalate, especially due to climate change. They are different from other natural hazards as they are mainly triggered by human interventions rather than natural triggers. The wildfire risk management is a complex process with many uncertainties in the assessment, fire behavior and spread modelling, and decision making. To predict wildfires, sophisticated temporal geospatial methods are required. This paper develops a wildfire probability prediction method considering the capabilities of Bayesian networks and fuzzy logic that can handle uncertainties and update probabilities in response to the availability of new data. The model takes into account the data from a geographic information system (GIS) for a specific area at micro level to estimate the wildfire probability and is able to update the probability due to any planned or unplanned changes in the area. Therefore, the proposed method can feed to future macro and micro risk-based decision-making situations in wildfire prone areas. The method is evaluated through a sensitivity analysis and its performance is investigated through a case study in New South Wales (NSW), Australia.**

Keywords—wildfire prediction, Bayesian networks, uncertainty, fuzzy systems

I. INTRODUCTION

Wildfires or forest fires, also called bushfires in Australia, are significant threats to lives and properties and are likely to be exacerbated by ongoing climate change factors of increased temperatures, variability in moisture conditions and forest clearance. In 2018, wildfires heavily affected Sweden, UK, Ireland, Finland and Latvia - all countries in which such fires had not been a concern in previous years. In 2017 wildfires burned over 1.2 million hectares of natural lands in the EU and killed 127 fire fighters and civilians [1]. Australia, which has a Mediterranean climate in some part of it, is also fire-prone, experiencing wildfires throughout the year in different parts of the country as shown in Fig. 1. Bushfires have a long history in Australia, going back 65 million years and indigenous Australians developed the first human fire management of bushland up to 60,000 years ago. Over the period of 1901-2011, 260 bushfires have been recorded with a total of 825 civilian and firefighter fatalities [2]. Each year, between 3% and 10% of Australia's land area burns with an estimated damage of $77 million per year, a loss calculated without considering the timber plantation cost [3]. This shows the importance and the urgency for serious investment in the prevention of wildfires.

Wildfires are hydro-geological events which are bound to escalate, especially due to climate change wherein every degree increase in warming leads to a 12% increase in lightning which is a major wildfire trigger, while the same level of warming requires 15% more precipitation to offset wildfire risk [4]. Wildfire, as with any fire, needs an ignition triangle consisting of oxygen, heat source and fuel, like trees, grasses, dry shrubbery and forest litter. Main natural sources of ignition are lightning, hot winds and even the sun. However, in the vast majority of cases, it is human intervention that creates fires, which, under extreme weather conditions, spread uncontrollably [1].
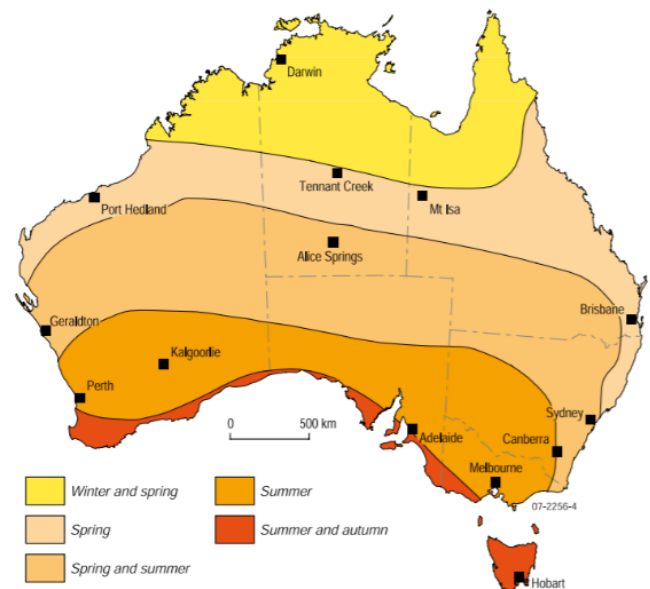

Fig. 1. Fire seasons across Australia [3]

Wildfires are more complicated than other fires in that, as they grow large enough, they create their own weather and increase their speed [4]. They are even able to create worse consequences with their expansion into neighbouring fire-prone areas, possibly creating a domino effect in which a primary wildfire spreads to (secondary) adjacent residential,

commercial, or industrial areas [5]. In May 2016, forest fires burned part of Fort McMurray in Alberta, Canada and moved towards the oil sands operation facilities north of the city. Although the fire did not reach the highly inflammable operations due to a fortuitous shift in wind direction, it resulted in a 40% drop in oil production due to the shutdown for safety, resulting in a loss of $760 million [4].

Managing wildfire risk involves analyzing both exposure and effects (i.e., likelihood and magnitude of potential beneficial/detrimental effects), then developing appropriate management responses to reduce exposure and/or mitigate adverse effects. Within this process, there are a number of linguistic, variability, knowledge and decision uncertainties [6]. The linguistic uncertainty refers to multiple definitions of wildfire risk. Variability uncertainties represent the frequency and spatial pattern of ignition locations, or the weather conditions driving extreme wildfire behaviour. The uncertainty in the scientific understanding of wildfire occurrence process and how it is modelled is considered as knowledge uncertainty. This includes the nature and quality of the data used to inform those models and the propagated uncertainty in model outputs. The last class of uncertainties in wildfire risk management denotes decision uncertainty in which preferences are elicited and social cost/benefit analyses are done for wildfire risk reduction decision making [6].

The aim of this paper is to develop a new wildfire prediction method considering the capabilities and characteristics of Bayesian networks (BNs) and fuzzy variables in which some of the uncertainties regarding variability and knowledge uncertainties are managed. BNs, also called Bayesian belief networks, are a mathematical graphical representation method that provides an opportunity to model the causal relationships among a set of variables, capture the weight of their influence on each other and update the posterior probability of variables given new evidence. The usage of fuzzy numbers helps with handling the vagueness of observable variables. In addition, the research methodology can be fed into a geographical information system (GIS) which facilitates capturing, storing, manipulating, analyzing, managing and presenting spatial or geographic data.

The rest of our paper is organized as follows: Section II reviews the literature and Section III presents the research methodology. Section IV shows the implementation and results. Section V concludes the paper and suggests future research directions.

II. LITERATURE REVIEW

A. Wildfire Risk Assessment

Fundamentally, a wildfire risk assessment as illustrated in Fig. 2 comprises the three main elements of likelihood, intensity and effects [7, 8]. According to this framework, wildfire risk is the expectation of loss or benefit that may occur to any number of social and ecological values affected by wildfire [7, 8].

Wildfire likelihood is viewed as either ignition probability or burn probability, depending on objectives. The ignition probability is typically estimated by statistical models relying upon wildfire occurrence data, while simulation techniques are used for burn probability estimation. For instance, the fuel management studies rely on burn probabilities while the ignition possibility is used in finding wildfire triggering hotspots [8]. Globally, many wildfire ignitions (around 80%) are due to human activities. Of this figure it is believed 90% happen in Mediterranean climate countries. Therefore, many studies on ignition probability estimation incorporate anthropogenic variables. These include housing density, distance from transportation routes and facilities, density of agricultural and livestock activities and other land use [9]. For the naturally-induced ignitions, physical environmental variables such as fuel moisture, relative humidity and temperature are important [8]. For large wildfires spreading over long distances, likelihood is better represented by burn probability. Thus, spatial ignition patterns may accurately reflect wildfire likelihood where fires are small (e.g. 1–10 ha); but where fires are larger, the estimates of likelihood need to account for wildfire spread from distant ignitions [8]. In Australia, the PHOENIX spread model is used for bushfire risk management. This model takes into account the fuel, weather and topographic conditions as a bushfire grows and moves across the landscape [10].
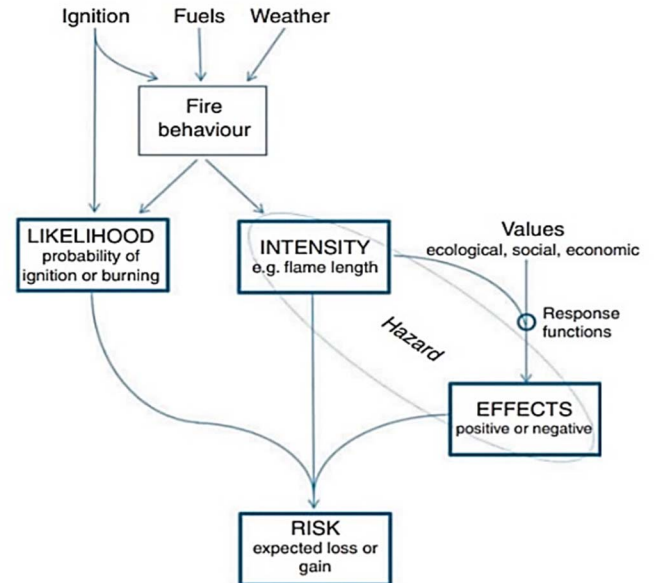

Fig. 2. A general wildfire risk framework [8]

Wildfire intensity is typically explained by flame length. The effects analysis explores the potential consequences of varying levels of highly valued resources and asset exposure to different wildfire intensities. If an area is modelled as a grid, the risk of wildfire in each cell i can be expressed as:

$$R_i = \sum_{j=1}^{M} \sum_{k=1}^{K} P(I_i) P(B_j|I_i) F(\mathrm{H}_k|B_j) \qquad (1)$$

where $P(I_i)$ is the ignition probability in cell i, $P(B_j|I_i)$ is the burn probability or the probability of the ignition evolving as a wildfire given intensity j and $F(\mathrm{H}_k|B_j)$ is the response function of highly valued resources and assets k to the wildfire of intensity j [11]. Therefore, the total risk for a specific area can be calculated as:

$$R_T = \sum_{i=1}^{N} \sum_{j=1}^{M} \sum_{k=1}^{K} P(I_i) P(B_j|I_i) F(\mathrm{H}_k|B_j) \qquad (2)$$

where N is the total number of ignitable cells. This calculation is static and does not include the temporal dynamics of wildfire risk.

During the last decade, there have been many attempts to formalize wildfire risk modelling and handling uncertainties in wildfire risk management. A review of advances in risk analysis for wildfire management can be found in [8] and a survey of methods in wildfire-induced industrial accidents can be seen in [12]. The use of BNs in these areas has received more attention in recent years. A BN model representing the factors influencing wildfire occurrence in Swaziland was presented in [9] and a dynamic BN model for wildland urban interfaces was developed in [11].

B. Bayesian Networks

A BN is composed of arcs and nodes. The nodes are the variables and the arcs symbolize the relations between those variables. In this way, BN could be an approach to see the cause and effect relations within a system. For each node, there is a conditional probability table (CPT) which represents how the nodes can influence each other according to the links. The principal feature of the BN is that it allows the possibility to model and think logically about uncertainties.

BN presents the joint probability distribution P(X) of variables $U = \{A_1, \dots, A_n\}$ based on the conditional independence resulting from the d-separation [13]:

$$P(X) = \prod_{i=1}^{n} P(X_i \mid Pa(X_i)) \tag{3}$$

where $Pa(X_i)$ is the parent set of X_i for any $i = 1, \dots, n$. If $Pa(X_i)$ is an empty set, then X_i is the root node and $P(X_i \mid P_a(X_i)) = P(X_i)$ denotes its prior probability [14].

Bayesian networks take advantage of Bayes's theorem to update the prior occurrence probability of events given new information. This new information or evidence usually can become available during a system's operation or the life of a process, including the occurrence or non-occurrence of an incident:

$$P(X|E) = \frac{P(X,E)}{P(E)} = \frac{P(X,E)}{\sum_x P(X,E)} \tag{4}$$

This equation is used for predicting or updating probability in a given BN.

C. Fuzzy Bayesian Networks

Now, suppose $X = \{X_1, X_2, \dots, X_n\}$ be the set of variables in a BN. If the variable X_i is a continuous variable then it can be transformed into a fuzzy random variable W_i. The corresponding set U_i can be used to map the variable X_i to fuzzy states:

$$U_i = \{\hat{X}_{i1}, \hat{X}_{i2}, \dots, \hat{X}_{im}\} \tag{5}$$

where $\hat{X}_{ij}$ is the j-th fuzzy state and m denotes the number of the fuzzy states, and fuzzy state $\hat{X}_{ij}$ can be defined as:

$$\hat{X}_{ij} = \{\mu_{\hat{X}_{ij}}(x) | x \in \boldsymbol{X}_i\} \tag{6}$$

where $\mu_{\hat{X}_{ij}}(x)$ is the membership function of fuzzy state $\hat{X}_{ij}$, $\boldsymbol{X}_i$ is the frame of X_i, and x denotes the value of variable X_i. For a continuous variable X_i, its condition probability in the BN with its parent can be replaced by $P(W_i|Pa(W_i))$:

$$P(X_i|Pa(X_i)) \rightarrow P(W_i|Pa(W_i)) \tag{7}$$

where W_i is the corresponding fuzzy random variable defined by X_i. For a node without parents, soft evidence is equivalent to modify its prior probability; otherwise, soft evidence on a variable X_i is represented by a conditional probability vector $P(X_i = x|H_i)$ for $i = 1,2,\dots,m$, where H_i denotes the hypothesis that the true state is the i-th state. To simplify the inference process for continuous variables, consider the fuzzy random variable W_i with states $\{\hat{X}_{i1}, \hat{X}_{i2}, \dots, \hat{X}_{im}\}$. Define $H_j, j = 1,2, \dots, m$, as the hypothesis that W_i is in state $\hat{X}_{ij}$. The results of fuzzy function member $\mu_{\hat{X}_{ij}}(x), j = 1,2, \dots, m$ form the soft evidence vector:

$$e = \{\mu_{\hat{X}_{i1}}(x), \mu_{\hat{X}_{i2}}(x), \dots, \mu_{\hat{X}_{im}}(x)\} \tag{8}$$

$\mu_{\hat{X}_{ij}}(x)$ is considered to be approximately equivalent to the condition probability $P\left(\mu_{\hat{X}_{ij}} \middle| X_i = x\right)$. Then the soft evidence vector can be defined as:

$$e = \{P(W_i = 1|H_1), P(W_i = 1|H_2), \dots, P(W_i = 1|H_m)\} \tag{9}$$

where $P(W_i = 1|H_j)$ represents that the observed value of W_i is "1" if the state is $\hat{X}_{ij}$ which is indeed the probability $P(\mu_{\hat{X}_{ij}}|X_i = x)$ [15].

III. RESEARCH METHODOLOGY

The research methodology follows a general methodology for constructing BN models. As shown in Fig.3, it includes seven major steps.

Main variables are identified through the literature review and experts' knowledge. Usually, a number of topographic, climatic, fuel and human variables are relied upon. Based on the variables identified, data are collected through several sources, including remote sensing datasets and morphological records. The BN structure is constructed based on the experts' knowledge or can be learned from the data or a combination of both. The CTPs are learned from the data too; however, experts' knowledge can also be incorporated. The BN model is evaluated through a sensitivity analysis which here involves the study of the effect of small changes in local parameters on the target node, meaning the important parameters and structure of a complex system can be evaluated [14]. For continuous variables, membership functions are determined by the parametric (trapezoidal/triangular) functions which are good enough to capture the vagueness of variable states. The BN model provides the wildfire probability for each cell of the area under study. The conditional probabilities of the form $P(wildfire|event)$ are calculated in predictive analysis which shows the occurrence probability of wildfire given the states of the primary events. In updating the analysis, the conditional probabilities of $P(event|wildfire)$ can be assessed, indicating the probability

of a particular event in a certain state or different possible states given the occurrence of the wildfire.
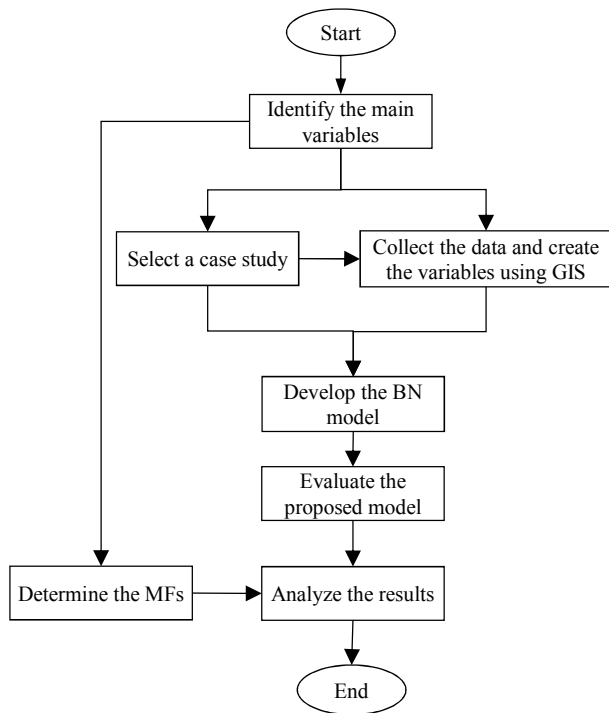

Fig. 3. The research methodology

IV. IMPLEMENTATION AND RESULTS

A. Study Area

The Northern beaches is a district of NSW, Australia, where the population of 253,000 makes it the third most populous local government area in the Sydney region. It occupies approximately 278 km2 of the area from 151.36 East to 33.58 South and includes a variety of land cover such as natural conservation forest, urban, horticultural and native vegetation (Fig. 3). This zone has experienced several bushfires since 2015, fires that initiated in forest areas and spread to rural regions and urban infrastructure. Fig. 4 shows the region and Fig. 5 shows the location of the case study in NSW.

B. Data Collection

In this study, we used the meteorological data of the Australian government's Bureau of Meteorology (http://www.bom.gov.au) from 22 nearby gauging stations between 2016-2019. Data included the means recorded for annual temperature, annual rainfall, annual humidity and annual wind speed. The land cover and topographical data were both derived from the NSW government database centre (https://data.nsw.gov.au/) in vector and raster formats, respectively, with 5-meter spatial resolution.
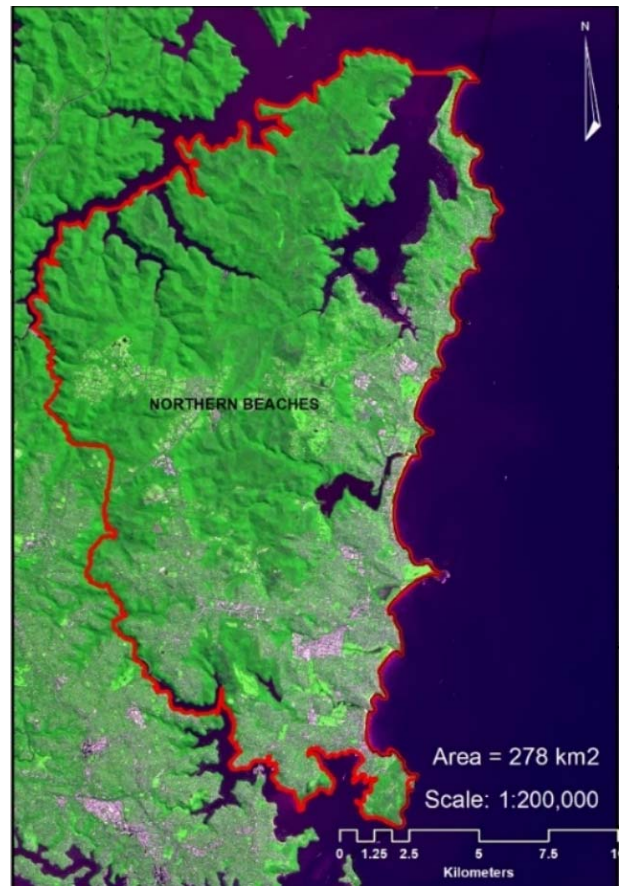

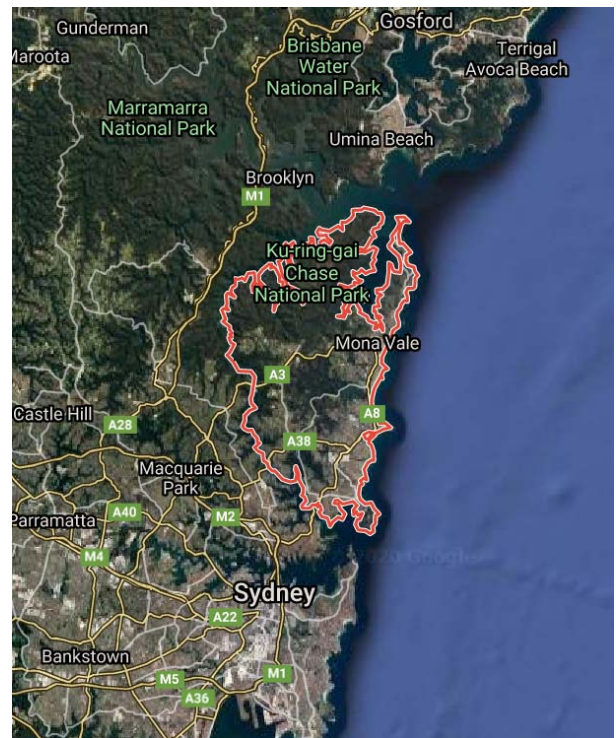Fig. 4. The location of Northern Beaches region

Fig. 5. The location of case study in NSW, Australia

C. BN Model

Nine independent variables, within four categories are considered. The topographic variables include elevation, slope and aspect. The climatic variables comprise mean temperature, moisture, and humidity. The fuel category has one variable–land cover. The human factor category consists of the distance to roads. An additional main variable is added to the BN model to represent the wildfire probability. Table I shows the variables and their types and states.

The BN model constructed based on expert knowledge is presented in Fig. 6. The CPTs are learned from the data using the EM algorithm in GeNIe platform (www.bayesfusion.com). The EM algorithm was proposed by Dempster et al. (1977) and provides an iterative procedure for maximum posteriori estimation in the case of incomplete data [16].

TABLE I. VARIABLES' PROPERTIES

Name	Type	States
Wildfire probability (WP)	Fuzzy	Low, Medium, High
Elevation (EL)	Fuzzy	Low, Medium, High
Slope (SL)	Fuzzy	Gentle, Moderate, Steep
Aspect (AS)	Discrete	North, East, West, South
Temperature (TE)	Fuzzy	Low, Medium, High
Moisture (MO)	Fuzzy	Low, Medium, High
Humidity (HU)	Fuzzy	Low, Medium, High
Land cover (LC)	Discrete	Natural conservation Water bodies Native vegetation Services and utilities Transportation Residential urban
Distance to road (DR)	Fuzzy	Close, Neither close nor far, Far

The inter-relationship among the variables and of each variable with WP was assessed using the BN model. According to the BN architecture, three variables are shown as direct contributing factors in wildfire probability – LC, EL, and MO. Other parameters, however, influence the WP through these three as shown in Fig. 6.

D. BN Model Evaluation

To ensure the BN model demonstrates acceptable behaviour, the sensitivity analysis is conducted, whereby the influence of variation in the model inputs on its outputs is systematically investigated. The inputs can be the BN parameters (i.e., conditional probabilities) or evidence (i.e. information about the states of nodes). Therefore, it is determined whether a variable is sensitive or insensitive to the changes in other variables.

The GeNIe software supports one-dimensional sensitivity analysis. The results are presented as bar charts in Fig. 7 which show the most sensitive nodes for WP in the occurring state in the absence of evidence. The red/green bars show the positive derivative values and the green/red bars represent the negative

derivatives. WP in an occurring state is most sensitive to SL, followed by MO and LC. As can be seen, after making a small change of 10% of its current value, the posterior output changes are not significant, ranging from 0.019 to 0.023. The network's structure and the prior probability values therefore seem fine.
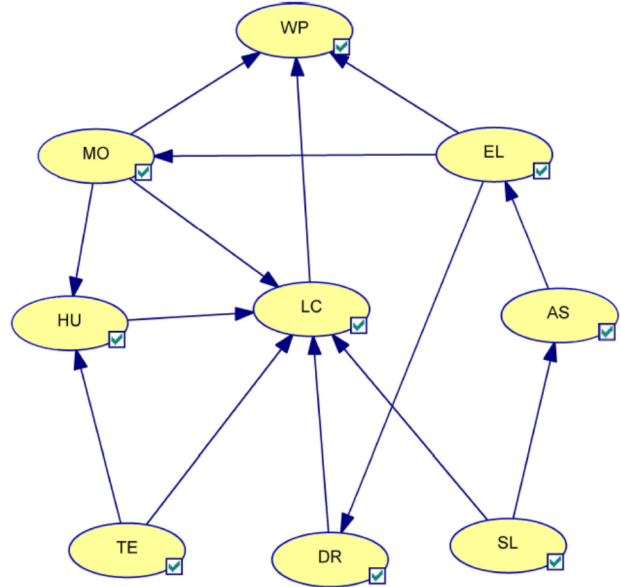

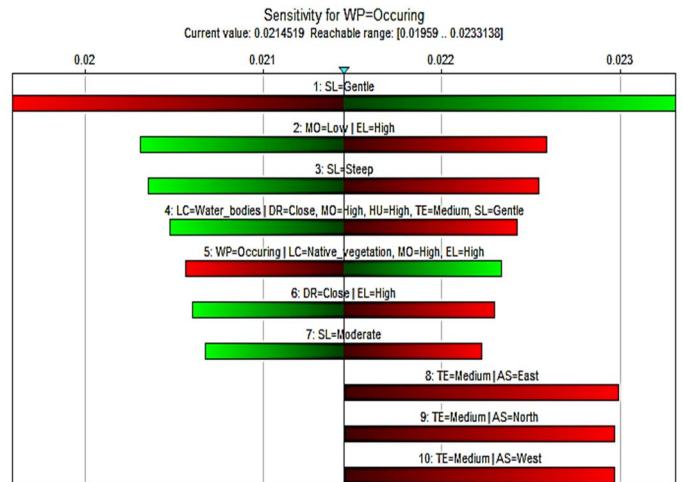Fig. 6. The BN model developed for the case study

Fig. 7. Sensitivity analysis for WP in occuring state

E. Variable Fuzzification

To analyze the results for each cell in the grid, first a fuzzification process is performed. Fig. 8 shows the membership functions of WP and DR variables, the remaining membership functions omitted due to limited space. As can be seen, a

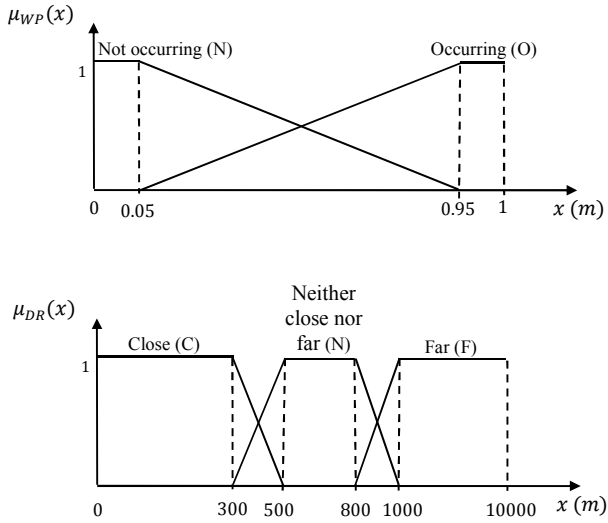combination of triangular and trapezoidal fuzzy numbers is used to increase the sensitivity in some bounds.

Fig. 8. WP and DR memebership functions

F. Results Analysis

The BN model is analyzed by GeNIe academic package. The wildfire initial probability of occurrence is 0.02 for the region. The highest prior probability of other variables and relevant states are summarized in Table II. The land cover, distance to roads and aspect are the most significant parameters since one of their states gained more than 59% of probability to WP. Since most of the study area falls into the medium state of TE, where the temperature remains unchanged almost for the entire area, it would not actively contribute into WP, even though the medium state of TE achieved 98% probability. Having said that, it would be considered as an influential variable when the model is extended to incorporate state and national data.

TABLE II.   PRIOR PROBABILITIES

Variable	State with highest probability	Probability
WP	Not occurring	0.98
EL	High	0.41
SL	Steep	0.36
AS	North	0.29
TE	Medium	0.98
MO	Low	0.59
HU	Moderate	0.51
LC	Natural conservation	0.67
DR	Close	0.62

For each cell, if the properties through the fuzzification process are entered into the model, the probability is updated as posterior probability for that cell. For example, for a cell in Table III, the wildfire occurrence is calculated based on the membership functions and the results are presented by relevant linguistic terms as in Table III.

TABLE III.   WILDFIRE PREDICTION

Input/output variables		Cell
NO		653915
LC		1
DR		75
AS		207.68
SL		11.35
EL		156.97
TE		22.99
HU		66.6
MO		885.1
WP	Occurring	0.04
	Not occurring	0.96

The BN model incorporates qualitative and quantitative data in different forms. The model which is based on the probability theory widely used to quantify uncertainties, facilitates handling some uncertainties in wildfire risk management. It includes the main contributing factors and the uncertainties associated with these factors and parameters are propagated through the network to the final wildfire probability prediction. Therefore, it provides a coherent framework for representing and reasoning with uncertainty.

V. CONCLUSION AND FUTURE WORKS

In the last decade, severe wildfires have affected and continue to devastate many areas worldwide. In only 2019, bushfires across Australia burned an estimated 18.6 million hectares, destroying over 5,900 buildings, killing at least 34 people and one billion animals. Therefore, wildfire risk management needs more attention and reliable practices. This very complex process involves many uncertainties, from its definition, assessment, and scenario modelling, to eventual decision making. This paper develops a BN-based model for wildfire occurrence prediction considering three types of effective parameters comprising topographic, morphological and human variables which were generated by GIS techniques. Fuzzy soft evidence is then used from each cell of the grid to estimate wildfire occurrence. In this study, we reduced the uncertainty of WP by integrating BNs and fuzzy systems. This method of wildfire risk management can improve our understanding of the uncertainties in variability, knowledge and decision making. The model is implemented in a case study of Northern Beaches of NSW, Australia and major implications are discussed.

The future work of this research is to calculate the risk levels for each cell of the grid, considering the highly valued resources and assets involved. In addition, more contributing factors are to be discovered.

REFERENCES

[1] J. San-Miguel-Ayanz *et al.*, "Forest Fires in Europe, Middle East and North Africa," 2018.

[2] R. Blanchi, J. Leonard, K. Haynes, K. Opie, M. James, and F. D. de Oliveira, "Environmental circumstances surrounding bushfire fatalities in Australia 1901–2011", *Environmental Science & Policy,* vol. 37, pp. 192-203, 2014.

[3] C. Sullivan, "Natural Hazards in Australia: Identifying Risk Analysis Requirements [Book Review]", *Australian Journal of Emergency Management, The,* vol. 23, no. 2, p. 74, 2008.

[4] N. Khakzad, M. Dadashzadeh, and G. Reniers, "Quantitative assessment of wildfire risk in oil facilities", *Journal of environmental management,* vol. 223, pp. 433-443, 2018.

[5] M. Naderpour and N. Khakzad, "Texas LPG fire: Domino effects triggered by natural hazards", *Process Safety and Environmental Protection,* vol. 116, pp. 354-364, 2018.

[6] M. P. Thompson and D. E. Calkin, "Uncertainty and risk in wildland fire management: a review", *Journal of environmental management,* vol. 92, no. 8, pp. 1895-1909, 2011.

[7] M. P. Thompson *et al.*, "Development and application of a geospatial wildfire exposure and risk calculation tool", *Environmental Modelling & Software,* vol. 63, pp. 61-72, 2015.

[8] C. Miller and A. A. Ager, "A review of recent advances in risk analysis for wildfire management", *International journal of wildland fire,* vol. 22, no. 1, pp. 1-14, 2013.

[9] W. M. Dlamini, "A Bayesian belief network analysis of factors influencing wildfire occurrence in Swaziland", *Environmental Modelling & Software,* vol. 25, no. 2, pp. 199-208, 2010.

[10] K. Tolhurst, B. Shields, and D. Chong, "Phoenix: development and application of a bushfire risk management tool", *Australian journal of emergency management,* vol. 23, no. 4, pp. 47-54, 2008.

[11] N. Khakzad, "Modeling wildfire spread in wildland-industrial interfaces using dynamic Bayesian network", *Reliability Engineering & System Safety,* vol. 189, 2019.

[12] M. Naderpour, H. M. Rizeei, N. Khakzad, and B. Pradhan, "Forest fire induced Natech risk assessment: A survey of geospatial technologies", *Reliability Engineering & System Safety,* vol. 191, p. 106558, 2019.

[13] J. Pearl, *Probabilistic reasoning in intelligent systems: networks of plausible inference*. San Francisco, California: Morgan Kaufmann Publishers , INC., 1988.

[14] M. Naderpour, J. Lu, and G. Zhang, "An intelligent situation awareness support system for safety-critical environments", *Decision Support Systems,* vol. 59, pp. 325-340, 2014.

[15] H. Chai and B. Wang, "A hierarchical situation assessment model based on fuzzy Bayesian network", in *Artificial Intelligence and Computational Intelligence*, vol. 7003, H. Deng, D. Miao, J. Lei, and F. Wang Eds., (Lecture Notes in Computer Science. Berlin Heidelberg: Springer-Verlag, 2011, pp. 444-454.

[16] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm", *Journal of the Royal Statistical Society: Series B (Methodological),* vol. 39, no. 1, pp. 1-22, 1977.