Knowledge-driven Development of Causal Model for Predicting Fire Hotspot Escalation in Central Kalimantan Peatland

Ariesta Lestari

Bachelor of Information Technology (Petra Christian University)
Master of Computer Science (Gadjah Mada University)

A thesis submitted for the degree of Doctor of Philosophy at

Monash University in 2019

Faculty of Information Technology

Copyright notice

Contents

List of Tables

List of Figures

Knowledge-driven Development of Causal Model for Predicting Fire Hotspot Escalation in Central Kalimantan Peatlands

Ariesta Lestari, M.Cs, S.Kom
ales4@student.monash.edu
Monash University, 2019

Supervisor: Dr. Grace Rumantir
Associate Supervisor: Prof. Nigel Tapper

Abstract

Peatland wildfires frequently occurring in the province of Central Kalimantan, Indonesia is a damaging environmental problem on a global scale. These fires have not only destroyed millions of hectares of Indonesian forest, but also produced haze and released carbon into the atmosphere, causing economic and health problems and contributing to the global greenhouse gas emission problems. As has been advocated by various forest fire experts, to alleviate the occurrence of peatland wildfires, it is important to ensure that hotspots do not escalate into wildfires in the first place. The behaviours of hotspot escalations could be learned from data from past fires to build a model to automate prediction of future possible escalations, if sufficient amount of such data existed.

Research in modelling forest fires prediction system mostly depends on the historical data. There is currently insufficient historical data to build a data-driven model of fire escalation in peatlands. Instead, for peatland fires in Central Kalimantan, experts have written a lot about the nature of the fires in qualitative narrative format in the literatures. This collective knowledge of domain experts has the potential to be utilised for building a quantitative predictive model, a research area that in general has not been explored much and has never been explored in the domain of peatland fires in Indonesia.

This thesis tries to address the gap between data science and related forest fire science by proposing a workflow to quantify the causal relationship amongst factors contributing to the escalation of hotspots into peatland fires. This repeatable workflow incorporates information from the literatures and knowledge elicited directly from experts through surveys. The aim is to identify the factors contributing to the escalation of hotspots into peatland fires and determine their causal relationships in a probabilistic manner. Using this workflow, a probabilistic graphical causal model in the form of a Subjective Bayesian Network has been developed and its performance evaluated.

The performance of the causal model is evaluated in two ways by (1) using a small set of available historical hotspot escalation data that has been physically checked on the ground and (2) comparing it against the performance of the implementation of the guidelines of how to determine hotspot escalation published by the Indonesian Government's National Institute of Aeronautics and Space (LAPAN). The causal model performs comparably well against the implementation of the LAPAN model. Analyses on some individual known past hotspots demonstrate that the causal model can quantitatively explain the relationships amongst the climatic conditions, peat ecology and societal vulnerabilities to probabilistically determine whether or not a hotspot would escalate into a wildfire.

The predictions of the causal model yield escalation probabilities between 45.8% and 62.8% with median/mean 50%/52.3%. The narrow minimum-maximum probability range is a reflection of the fact that in building the model, the experts have been cautious or not been decisive enough in assigning the conditional probabilities of the causal relationships between the variables. This may be a reflection of uncertainties that (1) the existing factors may not have been defined well or (2) there may be other factors that have not been incorporated in the model. This can also explain why the societal vulnerabilities that have been reported in antropology literatures on Indonesian peatland fires as contributing factors to the anthropogenic nature of peatland fires, do not feature as determining factors in the predictions of the causal model. The median/mean values of the predictions show bias towards escalation, which may reflect bias in the minds of the experts due to the non-neutral labels of the model variables.

This thesis demonstrates that a useful automated predictive hotspot escalation model can be developed purely using a knowledge driven approach by engaging experts through workshops and surveys. Further rounds of knowledge elicitation process upon presenting the lessons-learnt from this first round of surveys to the experts would help refine the model and hence improve its performance.

Declaration

This thesis contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

Signature: [redacted]

Print Name: Ariesta Lestari

Date: 13 August 2019

Vita

Publications arising from this thesis include:

Lestari, A., Rumantir, G., Saharjo, B., Usup, A., Graham, L., Tapper, N., Vayda, A P., Yulianti, N., Teguh, Rony (2018)
Analysing Causal Factors of Peatland Wildfires: A Knowledge-based Approach. In *Pasific Asia Conference on Information Systems (PACIS)*. Yokohama, Japan.
Lestari, A., Rumantir, G., Tapper, N. (2016)
A Saptio-temporal Analysis of Forest Fires in Central Kalimantan, Indonesia. In *Pasific Asia Conference on Information Systems (PACIS)*. Chiayi, Taiwan.
Lestari, A., Rumantir, G., Tapper, N. (2015)
Deteksi Dini Kebakaran Hutan dan Lahan di Kalimantan Tengah | Early Detection for Forest Fires in Central Kalimantan, Indonesia. In *Jurnal Dialog Penanggulangan Bencana. Volume 6, No. 2, November 2015.*

Presentations arising from this thesis include:

Lestari, A., Rumantir, G., Saharjo, B., Usup, A., Graham, L., Tapper, N., Vayda, A P., Yulianti, N., Teguh, Rony (2018)
A Knowledge-based Wild Fire Prediction System for Peatlands. Poster and paper presented at *Australasian Computer Science Week*. Brisbane, Australia.
Lestari, A., Rumantir, G., Tapper, N. (2016)
Knowledge-based approach for developing Forest Fire Prediction System. Paper presented at *Annual Conference of the Australasian Bayesian Network Modelling Society (ABNMS)*. Perth, Australia.
Lestari, A. (2016)
Save the Indonesian forest: A knowledge-based fire prediction system. Poster

presented at the *Monash FIT Innovation Showcase*. 18 October 2016, Melbourne, Australia.

Lestari, A., Rumantir, G., Tapper, N. (2015)

An Intelligent Forecasting System for Forest Fires in Central Kalimantan. Paper presented at *Doctoral Consortium Australasian Conference on Information System (ACIS)*. Adelaide, Australia.

Permanent Address: Faculty of Information Technology
Monash University
Australia

This thesis was typeset with LaTeX 2ε[1] by the author.

[1] LaTeX 2ε is an extension of LaTeX. LaTeX is a collection of macros for TeX. TeX is a trademark of the American Mathematical Society. The macros used in formatting this thesis were written by Glenn Maughan and modified by Dean Thompson, David Squire and Tian Yu Goh of Monash University.

Acknowledgments

The journey of my PhD was not an easy road. But, I am grateful to have a community, a support system that always encourages me, especially in the difficult times.

My earnest gratitude goes to my supervisors, Dr. Grace Rumantir and Prof. Nigel Tapper, whom I treat with the most respect. I would like to thank both of you for always encouraging my research, helping me grow as a researcher, and for the friendship and sympathy. It was a great privilege and honor to work and study under their guidance.

I express my gratitude to all the fire experts involved in this research, who have voluntarily shared their time, ideas and experiences. For the people whom I had the opportunity to interact with at the Caulfield School of IT. I am grateful to Prof Sue McKemmish, Prof Frada Burstein, Dr. Maria-Indrawan-Santiago, Prof Srini, and Dr. Julie Holden for their guidance, comments and advises. For all the admin people, Helen, Allison, Danette, Sid, Diana, and Rafiq who continuously help me throughout my PhD degree.

I also want express my gratitude to Australia Government through DFAT for funding my doctoral research through Australia Award Scholarship. I also acknowledge the Australian Government through the Australia-Indonesia Institute of the Department of Foreign Affairs and Trade for partially funded my data collection. I am also grateful for my institution, Palangkaraya University, who give permission in undertaking this doctoral degree.

This thesis has benefited from copyediting and proofreading work by Dr. Arvind Ratnaike, who has made this thesis more readable in its current form.

For friends and colleagues. All the member in my life group *Spencell*, thank you for always being supportive and keeping me in your prayers. I also would like to thank my fellow friends, particularly Yuri, Tennyson, Tian, and Prajwol, who always be around me for the past few years, making me laugh and making sure that I am well fed. For all the fellow students and alumni at Monash FIT, particularly Nandini, Liz, Megan, Peter, Greg, Sepher, Abdullah, Ying, Mozhdekh, Anindita, Haniz, Wud, Alexey, Ayesha, Ekjort, Vi, Dharsani, Lisa, Monisha. Thank you so much for the friendship. I also want to thank the team members and kids that I served in *Giggle*, and the Indonesian Dance Group, Lenggok Geni. Thank you for being my *valuable procrastinations*. For my housemates, Afentina, Daniel, Zara, Anya and Mally. Thank you for the friendship, the meals, the trips and for always kept the house clean. Special thanks to members of group *Caulfield Ceria*: Bu Misita, Dwi, Pak Iwan, Danny, Pak Hadi, Mas Irawan, Kadek, Mbak Utari. Thank you for all the advises and also the fun time.

I am extremely thankful for the support from my family in Indonesia. Words can not express how grateful I am to my mother, who always support my decision in pursuing higher education. I would also like to thank my sister and brother. Thank you for always supporting me in everything decision that I take, and encouraging me throughout this journey.

Finally, praises and thanks to my God, Jesus Christ, for His showers of blessing throughout this PhD journey. I have experienced your guidance and provision day by day and You are the one who let me finish my degree. Thank you Lord.

Ariesta Lestari

Monash University
August 2019

Abbreviations

ADB	Asian Development Bank
AHP	Analytic Hierarchy Process
ARI	Acute Respiratory Infection
BAPPENAS	Badan Perencanaan Pembangunan Nasional (*The Indonesian National Planning Agency*)
BKSDA	Balai Konservasi Sumber Daya Alam (*Nature Conservation Agency*)
BMKG	Badan Meteorologi, Klimatologi, dan Geofisika
BN	Bayesian Network
CFFDRS	Canadian Forest Fire Danger Rating System
CPC	Compatible Parental Configuration
CPT	Conditional Probability Table
DAG	Directed Acyclic Graph
DSR	Design Science Research
DSRM	Design Science Research Methodology
ENSO	El Niño-Southern Oscillation
FDR	Fire Danger Rating
FDRS	Fire Danger Rating System
FFDI	Forest Fire Danger Index
FFMC	Fire Fuel Moisture Code
FWI	Forest Fire Weather Index
GDP	Gross Domestic Product
GNP	Gross National Product
GWL	Ground Water Level
IDW	Inverse Distance Weighting
IPB	Institut Pertanian Bogor

IRI	International Research Institute for Climate and Society
KFCP	Kalimantan Forest and Climate Partnership
LAPAN	Lembaga Penerbangan dan Antariksa National (*National Institute of Aeronautics and Space*)
LDA	Latent Dirichlet Allocation
MODIS	Moderate Resolution Imaging Spectroradiometer
MRP	Mega Rice Project
NFDRS	National Fire Danger Rating System
NGO	Non-Government Organisation
NOAA - AVHRR	National Oceanic and Atmospheric Administration - Advanced Very High Resolution Radiometer
PMC	Peat Moisture Content
TRMM	Tropical Rainfall Measuring Mission
VEM	Variational Expectation Maximization

Chapter 1

Introduction

1.1 Motivation

Forest fires have become an increasingly serious environmental problem in many countries. During the massive forest fires in 1997-1998, countries in Southeast Asia and Latin America lost more than 20 million hectares of their tropical forest. The impact of the loss of tropical forest has threatened the existence of many varieties of indigenous plants and animals. The fires have changed and destroyed habitats of animals, making them lose their places to live and find food (Smith et al., 2000). The fires have affected the national or regional economic stability in many countries. For example, the 1998 fires in Brazil cost more than U.S $36 million in crop damages (Cochrane, 2010). In Indonesia, the 1997 fires have destroyed timber and agricultural crops. The damage is estimated at between U.S $3.6 billion and U.S $9.7 billion dollars (Tacconi, 2003).

However, the most threatening impact of forest fire is the loss of human life. The impact can be direct loss of life especially for the fire fighters and the forest residents (Whittaker et al., 2013), or indirect through the exposure to the air pollutants during or after the fires (Aditama, 2000; Kunii, 1998). During the fire incidents in 1997 and 1998, tens of millions of people were heavily exposed to the damaging inhalable or respirable air particles throughout Southeast Asia (Cochrane, 2010). Millions of people in Central America, North America, Africa, Australia and other parts of Asia were also exposed to smoke from these tropical fires. These people are now prone to severe respiratory and cardiovascular problems which lead to the increasing risk of mortality especially in children and the elderly (Aditama, 2000; Kunii, 1998). For example, during the 1997-1998 fires 527 people were officially reported dead in eight provinces that were affected by haze in Indonesia. However, based on unofficial reporting, Kunii (1998) estimated that the smoke caused by the 1997-1998 fires may have been responsible for nearly 31,000 deaths in Indonesia. Brazil also suffered from the haze caused by fires in its country in 1998. In Brazil, 700 deaths were reported due to respiratory problems (Cochrane, 2002). The health impact of forest fires goes beyond the immediate mortality statistic, because the long-term implications of unhealthy and dangerous air quality impacts the future generations. Efforts to minimise the impact of forest fire should become a major concern in these areas to provide a better quality of life for the future generations.

In Indonesia, fires that occurred in the peatland areas have seriously impacted human lives. These fires are not only difficult to extinguish, but also produce smoke that contain dangerous particles to human health (Adinugroho, 2005; Page and Hooijer, 2016). The impact of peatland fires was not only felt by Indonesia but by neighbouring countries such as Malaysia and Singapore. Therefore, the pressure from the global community has forced the Indonesian government to strengthen its efforts to reduce the risk of these fires (Djalante and Garschagen, 2017). A range of policies and regulations has been drawn up by the Indonesian government as part of the fire management system to tackle the peatland fire problems, however most of the implementations have failed or proved ineffective (Herawati and Santoso, 2011; Saharjo, 2016). Based on the experience of the previous fire incidents, it was found that the current regulations mostly emphasize fire control, suppression of the fires, and emergency response rather than the prevention of fire outbreaks (Adinugroho, 2005; Herawati and Santoso, 2011). If the fires are of low to moderate fire intensity, they can be suppressed easily. However, once a fire escapes from the suppression actions and continues spreading without control, it will be difficult to extinguish. It can be a costly action. Therefore, fire prevention is needed as an early activity in the fire management system.

In many different countries, implementing an early warning system or fire detection system is one way to help prevent the occurrence of forest fires and reduce the impact of fires. A fire danger rating is one of the modelling systems implemented across Australia, Europe, and North America as an early warning system for predicting fire occurrence. Australia has developed a fire danger rating system based on the McArthur model which takes into account fuel and the climatic condition. Based on the forecast from the fire danger rating, the Bureau of Meteorology issues fire warnings to the public. The fire warnings are followed up by land management agencies issuing Total Fire Bans (Hennessy et al., 2005). The Canadian Forest Fire Danger Rating System is used across Canada during the fire season. A daily forecast on the severity of fire weather conditions is issued daily for common standardised forest types. This system has been adopted by, or adapted to, a number of countries around the world such as New Zealand, Mexico, Portugal, Malaysia, and even

Indonesia (De Groot et al., 2007). However, the implementation of this fire danger rating system in Indonesia seems inefficient and unsuccessful, in part due to lack of recognition of the unique peatland fire characteristics. These characteristics include the characteristic of peat itself and human involvement. Researchers agree that a better understanding of the behaviour of peatland fires could provide more reliable information that can support in the decision making to prevent fire occurrences (Applegate et al., 2002; Dennis et al., 2005).

In Indonesia, a vast amount of knowledge about fire escalation on peatlands is available in a multitude of published literature and from human experts. This knowledge from literature and human experts can be used to understand the behaviour of peatland fires from the perspectives of anthropogenic fires and climate and ecology science. However, the modelling approach in the data science technique mostly relies on historical data rather than incorporating the domain knowledge. In Indonesia, the problem with insufficient historical data has become a challenge for studying the characteristics of these peatland wildfires under the perspective of data-driven approach. When the data from experimentation or observations are limited, the acquisition of human knowledge can supplement the information (Meyer and Booker, 2001). Since there is gap in incorporating the domain knowledge in peatland fire science and data science, this thesis aims to incorporate cross-disciplinary research for the knowledge-based approach. The combination of information in the literature and knowledge from human experts can be used to describe the behaviour of hotspot escalation into peatland fires in Central Kalimantan. A probabilistic graphical model using Bayesian Networks is developed to incorporate all the factors influencing the behaviour of peatland fire . Once the behaviour of fires is studied and can be explained, the escalation of hotspots into surface peatland fires can be predicted. This prediction is essential in the prevention of peatland fire occurrence in Central Kalimantan.

1.2 Research Background

1.2.1 Peatland Fires in Indonesia

Forest fires are not a new phenomenon in Indonesia. In the modern era evidence of forest fires in East Kalimantan is documented in the early 1980s (Goldammer and Seibert, 1990), when the fires destroyed 3.6 million hectares of land and forest. Since then, fires have happened almost every three years especially during the El-Nino phenomenon. In the period of 12 years, from 2000 to 2012, Indonesia lost more than 6 million hectares of forest cover due to forest fires (Margono et al., 2014)

Forest fires in Indonesia occur not only in dryland but also in wetlands such as peatland area (Chokkalingam and Suyanto, 2004). Typically peatland is covered with closed forest, permanently moist and has a high degree of fire resistance (Chokkalingam and Suyanto, 2004; Page and Hooijer, 2016). However, destructive human activities have made the peatland degraded and vulnerable to fire occurrence (Harrison et al., 2009; Miettinen and Liew, 2010). Fires happening in peatland areas are more difficult and challenging to suppress compared to fires occurring in dryland. These fires not only burn the surrounding area, but also can burn the underground peat, and continue to smoulder for days even weeks (Usup et al., 2004; Rein et al., 2008). These smoldering fires are not easy to identify and can only be extinguished naturally by heavy rain (Usup et al., 2004; Page and Hooijer, 2016).

Human activities have caused peatland areas to become degraded and prone to fire occurrences. The conversion of peatland areas to agriculture, industry, or settlements has changed the original nature of these areas. The use of fires in human daily activities has also become the primary cause of peatland fire occurrences. Activities such as the exploitation of natural resources, construction of canals, and showing the land ownership, are reliant on the use of fires (Dennis et al., 2005; Harrison et al., 2009). Using fire to prepare the land for agriculture and plantations has become the preferred method of land preparation because it is not only the cheapest and easiest method, but it also produces minerals which can be absorbed easily by plants (Saharjo, 1999; (Adinugroho, 2005).

Even though the peatland fires in Indonesia are mainly human-made, they are also influenced by a number of factors such as climatic conditions and physical conditions. The hot and dry conditions during the dry season compose one of the factors that increases the likelihood of fire. The risk of fires even increases during the El Niño phenomenon because the dry season is sometimes prolonged with extremely low rainfall and extremely high sunlight (Putra et al., 2011). The changes in the physical condition of peatland are mostly caused by the illegal logging, the conversion of the peatland for settlements, agriculture, and mining (Miettinen and Liew, 2010). These changes have degraded the peat, reduced its capability to absorb water and making it prone to fire occurrences (Cochrane et al., 1999).

Central Kalimantan is one province in Indonesia that has experienced extensive peatland fires over many years. This province has one of the largest tropical peat swamp forest areas world-wide. It remained pristine until the early 1990s(Rieley et al., 1996; Siegert et al., 2002). However, in 1996 the Indonesian government initiated a conversion project to turn a million hectares of peat swamp forest into areas for rice cultivation and transmigration (Notohadiprawiro, 1998). This project failed and left a million hectares of abandoned and degraded peatland (Muhamad and Rieley, 2002). This area became one of the major regions to have experienced extensive fires and brought severe damage to the environment(Siegert et al., 2002; Harrison et al., 2009; Hoscilo et al., 2011). The changes in the landscape of peat swamp forest due to the fires included changes in vegetation types. This has increased the likelihood of fires recurring with greater severity. Protecting undisturbed forest and preventing the recurrence of peatland fire should be a high priority in forest fire management systems.

1.2.2 Fire Management System

The purpose of forest fire control is to protect the forest from wildfire. The activities in fire control include preventing the occurrence of forest fire, suppressing small fires before they spread, and using fire for certain purposes on a limited scale (Adinugroho, 2005). The prevention of forest fire occurrence is one important component of the fire management system to reduce or minimize the number of fire incidents. This prevention system should be able to provide information, about the possibility of fire breaking out, to all the relevant stakeholders (Saharjo, 2016). Conventionally, such information has been provided by the fire authorities through direct observation on the ground . However, with the help of modern technology, it is possible to develop an information system to deliver early detection and prediction of the possibility of fire occurrence to the fire authorities and the local community.

The objective of early detection is to detect the occurrence of forest fires in the early stages and the exact location, even before the outbreak becomes established (Alkhatib, 2014; Mahdipour and Dadkhah, 2014). There are a number of detection and monitoring systems used by fire authorities. These include manual observation on the ground by patrolling the forest area or building monitoring towers, aerial and satellite monitoring systems, the use of optical networks and digital camera in the surveillance of the fires, and the deployment of wireless sensor network technology (Alkhatib, 2014).

In Indonesia, the most common method to detect and monitor fire occurrence is using hotspot information detected by satellites. The fire authority uses the hotspot information as an indicator of fire probability. However, since not all hotspots are fires, verification and confirmation of the hotspot occurrence is needed to determine which hotspot is a real forest fire (Saharjo, 2016). The verification of these hotspots is conducted through patrolling and ground observation by the fire authorities or the people in the nearest settlement. However, this method has been shown to be ineffective and an inefficient way due to a number of reasons such as limited access to the location of hotspots, lack of human resources to cover large areas, and the high costs involved in conducting ground observations (Alkhatib, 2014). Thus, Saharjo (2016) suggested that the early detection *system should be able to identify the escalation of hotspots into peatland fires and deliver accurate information about the location of a peatland fire once a hotspot is detected*. When the accurate location of potential fires is obtained, preventative actions can be taken immediately.

In the attempt to understand the escalation of hotspots into peatland fires and be able to build a detection system, it is essential to possess a comprehensive knowledge of peatland fire behaviour. Information about hotspot from the satellite together with climate and ecology data as the observed and experimental data is one of the fastest and easiest way to learn about the behaviour of peatland fires. However, when data from the experimentation or observation are limited, the acquisition of human knowledge can supplement the information (Meyer and Booker, 2001).

1.2.3 Knowledge Elicitation

Knowledge elicitation is the process used for the acquisition of knowledge by a system. This process includes all activities pertaining to the acquisition of knowledge from relevant sources, the analysis and interpretation of the knowledge, and presentation of the knowledge in a form that can be used for the system (Cordingley, 1989).

The relevant sources can be in the form of written material or knowledge from human experts. The analysis of written materials has been used to investigate the concepts, relations, and methods described in the literature (Payne et al., 2007). The written materials (henceforth termed as literature) are mostly published as textbooks, journal articles, or project reports. They provide experimental evidence, procedures, and the reasoning of findings that broadens the understanding of a worldwide phenomenon. Due to the rapid growth of literature, there is a need for techniques that help scientists to process the numerous sources of information (Hirschman et al., 2007). The technique should be able to manage automatic searching, extracting, organising, and indexing of information from a large collection of text resources. Based on research in machine learning, topic modeling is a method used for finding hidden patterns of words in document collections (Blei, 2012). The identification of patterns in word use leads to discovery of common topics that run through the documents, including the connection between those topics in the documents. The use of topic modeling can reduce the cost of labour and time spent in manual knowledge acquisition.

Human experts (henceforth termed experts) are people who have skills and knowledge of a specific domain that have been developed and built through complex experiences (Diaper, 1989). The knowledge from experts is important particularly when experimental and observational data, or validated computer models are not available (Meyer and Booker, 2001). The experts involved in this current thesis project consist of people that came from the Indonesian government, non-government organisations (NGOs), and university sectors. The experts might or might not live in the study area of Central Kalimantan.

Gathering knowledge from the experts can be done by directly questioning them about the problems of forest fires (Meyer and Booker, 2001). The elicitation process can be in the form of face-to-face interaction or written questionnaires. Face-to-face interactions can occur privately in one-to-one interviews, or in a group setting such as a focus group. The individual interview is one of the best methods to get detailed information and explanation of a problem. Furthermore, individual interviews reduce biases or influences that can arise through group-think (Meyer and Booker, 2001). However, there are certain drawbacks associated with individual interviews, such as the amount of time required to conduct the interviews, especially when there is a large number of experts.The use of a survey or questionnaire to gather expert knowledge is less expensive in terms of time required. It also allows the expert to complete the survey at their own pace (Meyer and Booker, 2001; Knol et al., 2010). Other drawbacks of individual interviews are different interpretations, different linguistic judgements, and the variance in explanations or understanding of a particular topic due to the diverse backgrounds of the experts (Knol et al., 2010). These problems can be minimised through a prior group discussion between the experts and data gatherers. Through a group discussion, a greater understanding of problems can be acquired and more accurate data generated (Meyer and Booker, 2001). However, group-think biases and scheduling meetings for experts are potential challenges of group discussions.

1.3 Knowledge Gap and Significance of The Work

Understanding the relationship between anthropogenic fire, climate, and ecological conditions is an early step in dealing with the complex problems of peatland fire in Indonesia. For peatland fire in Central Kalimantan, Indonesia, a massive amount of domain knowledge is mostly available in narratives and qualitative results (Applegate et al., 2002; Dennis et al., 2005; Adinugroho, 2005; Vayda, 2010; Cochrane, 2010). However, not much research is conducted to extract and utilise this available information for the modelling of causal relationships between anthropogenic factors, climatic conditions, and ecological parts of the peatland fires. The modelling of environmental problems such as forest fire mostly happens in the domain of data science. In this domain , the modelling most likely depends on historical data. The challenges in modelling an environmental problem is insufficient historical data and consistency of human expert (Eden, 1998). This challenge has become a barrier in obtaining reliable information about the behaviour of the fires (Finkel, 1996). Therefore, other available resources will be explored to understand the characteristics of the peatland fires.

This research addresses the gap between the data science and the related forest fire science. As shown in Figure 1.1, a cross-disciplinary collaboration is needed to model the hotspot escalation into peatland fire. This research includes a workflow designed to incorporate all the interdisciplinary perpectives presented in Figure 1.1 and to learn the causal relationships among contributing factors in peatland fire occurrences.

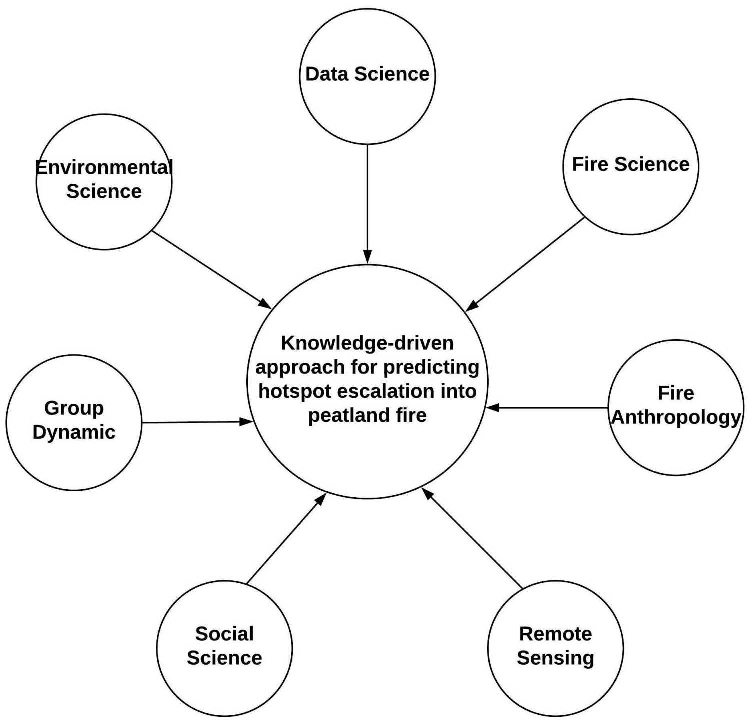


Figure 1.1: A cross-disciplinary collaboration in conducting a knowledge-driven approach for predicting hotspot escalation into peatland fire.

The knowledge gained on factors contributing to forest fire behaviour and on the data science technology for advanced data analytics facilitates the collaboration on the building of an automated causal model. This causal model can be used to draw inferences on the interactions between factors, and as an input to a fire monitoring system.

1.4 Research Objectives and Questions

The objective is to present the knowledge of the peatland fires' characteristics in a way that can easily be understood and used to support decision making to predict and prevent the occurrence of peatland fires. Using the proposed workflow, a causal model that can explain the behaviour of surface peatland fires in Central Kalimantan is constructed. The model is used to predict which hotspots may escalate into peatland fires. This is essential as it provides an early warning system for occurrence of peatland fires in Indonesia.

Three research questions have been formulated to address the challenges and aims discussed in the section above. These questions are presented as follows:

RQ.1: What can be learned from current approaches used in prediction models for escalation of peatland fires in Indonesia?

The first research question is about exploring the use of current approaches used by the fire authorities in predicting the fire occurrences. The aim is to understand the benefits and limitations of using such approaches. A comprehensive literature review is conducted to answer this research question. This question is broken down into two sub-questions.

> **RQ.1.1**: What are the current approaches in predicting peatland fires in Indonesia?
>
> This sub-research question leads to identification of current approaches and modelling methods implemented in Indonesia, for predicting forest and peatland fire occurrence .
>
> **RQ.1.2**: What method/approach might best be used to model the escalation of peatland fires in Indonesia?
>
> In this sub-research question, the challenges of the current methods/approaches to peatland fire escalation are examined (i.e. why they are not working). The idea is to identify the most suitable approach to understanding the behaviour of peatland fires. This information aids in developing a model that predicts the escalation of peatland fire.

RQ.2: In the absence of sufficient historical ground truth fire escalation data, can peatland fire escalation be modelled in a quantitative manner?

The second research question explores the use of a knowledge-based approach as a solution to the lack of appropriate models for understanding the complex behaviour of peatland fires and predicting escalation of hotspots to peatland fires. To answer this research question, the question is broken down into two sub-research questions.

RQ.2.1: How can information from the literature be extracted to identify contributing factors for peatland fire escalation?

The first sub-research question of Research Question 2 uses topic modeling as one of the methods in text mining to explore how the existing information in the literature can be extracted to define the factors contributing to the escalation of peatland fires. Information from the literature is used as there is lack of historical data that can be used in the analysis.

RQ.2.2: How can expert knowledge be incorporated to develop a comprehensive understanding of the characteristics of peatland fires and used to predict the escalation of peatland fires?

The purpose of this research question is to explore whether the use of expert knowledge as a knowledge source can contribute to the development of a peatland fire model. The knowledge elicited from the experts is quantified and presented in a graphical model using the Bayesian Network.

RQ.3: In the absence of sufficient historical ground truth fire escalation data and *the gold standard* model, how can the causal model be evaluated?

This third research question aims to present and test a causal model that can be used to predict the escalation of hotspots into peatland fires using the proposed development framework. The aim is to ensure that the developed causal model is able to capture the complexity of peatland fires in Indonesia by employing some model evaluation processes.

1.5 Contribution

This research is cross-disciplinary research that involves peatland fire science and data science technology along with elements of complex human behaviour. The contributions of this thesis to these disciplines within the research area are presented below:

1. A potentially repeatable workflow that quantifies the causal relationship amongst the factors contributing to escalation of hotspots into peatland fires. This generic workflow can be applied to solve real-life phenomena with complex and uncertainty problems. There are three aspects of the contribution in this workflow:

   (a) This repeatable workflow incorporates information from literature and knowledge from experts to identify the factors contributing to the escalation of peatland fires and determines the causal relationship. This workflow also can deal with the complex and uncertainty problem that occurs due to insufficient historical data.

(b) An improvement using topic modeling to the general process of the development of the causal model. Topic modeling can be used to help the elicitation. In the process of identifying the causal variables and incorporating experts' knowledge in the modeling process.

(c) The development of the causal model occurs in a multi-disciplinary domain with experts from different disciplines.

2. Topic modeling of the published literature can be used to automatically extract the influencing factors of hotspot escalation. The quality of the topic modeling result is proven to be complemetary to the expert's opinion.

3. Capturing the expert's thinking process about how hotspots escalate into peatland fires through the development of a causal model. The causal model has been able to bridge the gap between peatland fire science and data science. In addition to that, the causal model also can be used to support the decision making in preventing the escalation of hotspots into peatland fires.

1.6 Thesis Structure

To achieve the aim and address the research questions, this thesis requires 8 chapters.

Chapter 1: Introduction - This chapter begins by discussing the motivation in conducting this research project. The research background that briefly covers all the relevant topics related to this research project is also presented. Aims and research questions for the research are also discussed, followed by the significance of the project and its contribution to the body of knowledge and implementation.

Chapter 2: Literature Review - This chapter begins by discussing the history and current status of surface peatland fires in Central Kalimantan, Indonesia. This covers the impact of the peatland fire on biodiversity, the economy, and health, and the fire management system that is carried out by the government to minimise the impact of the fire. In the literature review, the use and construction of Bayesian Networks (BNs) as a problem solver is also discussed. Knowledge acquisition is introduced. This covers the method used to elicit expert knowledge and to combine multiple expert's answers. It also highlights pitfalls and biases that arise when eliciting knowledge from experts.

Chapter 3: Research Methodology - This chapter captures the details of the methods in this study. This thesis is framed using the Design Science Research (DSR) Methodology. A brief explanation of the DSR Methodology and the stages used in the methodology are described. The artefacts which form the contributions of this thesis are introduced followed by the process of how to evaluate the artefacts.

Chapter 4: Automated Identification of Causal Variables - This chapter provides a detailed methodology on how to elicit the initial variables to be used in the causal model from the literature using the topic modeling technique. This chapter also provides a detailed explanation of the chosen variables. It also compares the topic modeling results with the expert opinions. The output from this chapter is used in Chapter 5 to generate the initial structure of the causal model.

Chapter 5: Structure Development of the Causal Model - This chapter discusses the second stage of the workflow of causal model development. The process of eliciting expert knowledge to build the structure of the causal model is described. The implementation of the knowledge elicitation process in structuring the causal model for hotspot escalation into peatland fire is also presented.

Chapter 6: Parameterisation of the Causal Model - This chapter discusses the third stage of the workflow of causal model development. This chapter provides a detailed methodology for gathering the expert knowledge to elicit the conditional probability table (CPT) that is used to parameterise a BN. The challenges in eliciting expert knowledge through online survey are also covered in this chapter.

Chapter 7: Test Data Preparation, Test Result, and Analysis - This chapter provides information on the test data for the evaluation process. The performance of the causal model based on the result of test data is also presented. An analysis is carried out to investigate the performance of the causal model in handling real life problems.

Chapter 8. Conclusion and Future Work - This chapter presents the summary of the thesis and focuses on the contribution of the thesis to the theory and practice. This thesis closes with the recommendation of future directions for further studies.

Chapter 2

Literature Review

Introduction

This research project aims to investigate how the combination of information in the literature and knowledge from human experts can be used to describe the behaviour of a phenomenon. In this project, the characteristic of surface peatland fire escalation is described in the form of a graphical model using knowledge extraction from literature and human experts. Therefore, this chapter contains two parts: the first is a further explanation of peatland fires in Indonesia; the second covers the methods used in the knowledge elicitation process.

In the first part, the definition of peatland fires and the contributing factors together with the damage caused by these fires are presented. The explanation then continues with the importance of preventing the occurrence of peatland fires and the investigation of the current approaches, including the systems and tools that are already deployed by the Indonesian government. The limitations of those systems and recommendations on how to enhance the fire prevention system in Indonesia are also discussed in this chapter. The first part or this literature review aims to address Research Question 1 which concerns what we can learn from current approaches in prediction models for peatland fire in Indonesia. The second part of this literature review contains an explanation of the different methods that are used to gather information from the literature. This part also introduce the method to elicit the knowledge from human experts and how to develop the causal model.

This chapter is concluded with a summary of the problems in the current approaches implemented by Indonesian government and the knowledge elicitation process to learn the behaviour of this fire.

2.1 Peatland Fires

Peatland is a type of wetland that may be or may not be covered by vegetation. Soil in the peatland forms from remnants of organic material that have accumulated over a long period (Joosten and Clarke, 2002; Adinugroho, 2005; Cochrane, 2010). Peatland has a special characteristic that distinguishes it from other ecosystem types, which is the capability of the peat soil to absorb and store water (Joosten and Clarke, 2002). This characteristic enables the peat area to remain almost permanently moist and have a high degree of fire resistance (Page and Hooijer, 2016)

However, the stability of peatland as a medium for water storage has been threatened by human activities, such as logging activities, conversion to farmland and plantation, conversion to the settlement for the transmigration program, and excessive draining (Silvius and Diemont, 2007; Hayasaka et al., 2016). These destructive activities have made the peatland degraded and vulnerable to fire occurrence (Miettinen and Liew, 2010). Peatland fires do not only occur on surface of the peatland, but also can happen underground. Usup et al. (2004) defined surface peatland fire as fires that originate on the surface of the peat, usually from a slashed area and then spread out of control to bush vegetation or secondary peat forests. These fires can go downward, burn the underground peat, and spread to the peat dome and the area surrounding the tree roots. They are then known as underground peat fires or deep peat fires. Much research in the peatland fires field agrees that underground peat fires are more hazardous compared to surface fires. The fires in deep peat can persist for a long duration and are difficult to extinguish. The smoke resulting from deep peat fire contains dangerous particles (Adinugroho, 2005; Usup et al., 2004; Page and Hooijer, 2016; Turetsky et al., 2015; Rein et al., 2008). Furthermore, the fires in the surface peatland not only spread vertically but also spread downwards, and ignite the deep peat fires (Usup et al., 2004). If we can prevent the fires in the surface peatland from burning downwards, it will protect the underground peat from disastrous fires and prevent the hazard of the underground peat fire (Tishkov, 2010).

2.1.1 Factors Supporting the Peatland Fire Incidents

Peatland fires have occurred in Indonesia over several millennia, but they have become a more regular feature in recent years especially in Sumatra and Kalimantan. Research has found that 99.9% of the fire occurrence was ignited by human activities (Goldammer and Seibert, 1990; Adinugroho, 2005). The rapid change of land use, exacerbated by climatic variability, also has led to an increase in the fire frequency (Cochrane, 2010). In the next subsections, further explanations of the cause of peatland fires in Indonesia is provided.

a. Human activities

Fire has been part of the daily activities for the indigenous communities in Kalimantan and Sumatra. The Dayak communities, indigenous to Central Kalimantan, traditionally used fires as a tool to clear the land - a technique known as slash-and-burn (MacKinnon, 1996). They used this tool as it is a cheap, practical way to open up the land (Suyanto et al., 2009). Using fire to clear the land also can increase the soil fertility (Sorrensen, 2004) and help in eradicating pests (Kinseng, 2008). Despite the benefit of this technique in enhancing soil fertility, many researchers also believe that fire used for land clearing could turn into uncontrolled fires that escape to nearby forests and then become wildfires (Goldammer and Seibert, 1990; Varma, 2003; Sorrensen, 2004; Dennis et al., 2005). However, a few studies found that the land clearing activity in the Dayak communities does not have significant influence in the escalation of peatland fires. Vayda (1999) explained that during fieldwork in East Kalimantan, no evidence was found to show fire occurrence from the slash-and-burn activities conducted by the Dayak people. Suyanto et al. (2009) also argued that the Dayak farmers usually followed the traditional rules inherited from their elders on how to control the fire in the slash-and-burn technique. So, it is less likely that land clearing activity conducted by these farmers contributes to the occurrence of fires.

Peatland areas in Indonesia have also been selected for the development of an agriculture-based transmigration project. The Indonesian government initiated this project to improve the lives of poor and landless people, by offering them land and jobs in the "underpopulated" islands such as Kalimantan and Sumatra (Cochrane, 2010). Harrison et al. (2009) reported that in Sumatra, the probability of fires in the transmigration area are four times higher compared to other areas. Most of people living in the transmigration area are not the traditional shifting cultivators. They often use fires carelessly when clearing the land or trying to fertilize the land (Fearnside, 1997; Byron and Shepherd, 1998).

Not all peatland areas can be used for agriculture. Some peatland areas are poor in nutrients and flooded with water every year (Adinugroho, 2005; Nursyamsi et al., 2016). Thus, this condition has forced the people living in the peatland area to survive through hunting, fishing, illegal logging, and collecting non-timber forest products. Vayda (2011) studied the human activities that could trigger the fire occurrence. He explained that accidental fire ignition could come from the use of fire by the loggers or fire users. These people used fire for cooking, insect repulsion, or campfires; sometimes the fires were not fully extinguished when they left the site. One of the propositions suggested by Colfer (2002) mentioned that the carelessness of people in using fires had increased the danger of fire spreading. Therefore, many fires were found in the areas accessible by people. For example, fires were found along the rivers or along roads.

Fires were also often used as tools to reclaim the rights of land. In Sumatra, the smallholders sometimes burned the land to reassert claims over land that had previously been planted by timber or palm oil companies (Dennis et al., 2005). In West Kalimantan, villagers contested the right of timber plantations to operate in the border of the village. They simply allowed the fire to burn out of control to the company's newly planted trees, while their own plantation was protected. They also used fires to secure an area as the information recognition of their "private" property claims on land that was considered common property.

Although the human contribution has a crucial role in the ignition and spread of peatland fires, there are other factors influencing the fire occurrence such as climate condition and physical conditions.

b. Climatic conditions

The risk of a fire breaking out in the peatland area is highest during the dry season when the rainfall is extremely low. Much research has found a strong correlation between the low amount of rainfall and the increase of peat fire occurrence. Ceccato et al. (2010) used the rainfall anomalies to show the relationship between rainfall and the occurrence of hotspots. They found that in wetter conditions, the number of hotspots tends to be low; when the rainfall anomalies are below average, the number of hotspots increases. They believed that the use of rainfall anomalies delivers a better understanding of the influence of rainfall on hotspot occurrence compared to using the amount of monthly or daily rainfall. However, several studies used the daily rainfall data to investigate the relationship between daily rainfall and the occurrence of fire. The findings of these analyses also showed that during the dry period, when the amount of rainfall is below normal, the fire activities also increase (Putra et al., 2011; Yulianti and Hayasaka, 2013). Even though the researchers used different measurements or methods to find a correlation between rainfall and fire occurrence, all of them showed a similar finding. Rainfall has a significant influence on the occurrence of fire in Central Kalimantan. When the amount of precipitation drops below average, it creates dry conditions on the peatland area and increases the susceptibility of the peatland ignition(Putra et al., 2011).

Severe peatland fire occurrence in Borneo is driven by the prolonged drought that happens during the El Niño years. The evidence shows that during those years the dry season in Borneo could last for 4 to 5 months (Usup et al., 2004), while in the normal years the dry season only happens for 2-3 months (MacKinnon, 1996; Aldrian and Dwi Susanto, 2003). During the El Niño years, the dry season is much drier compared the dry season in the La Niña years (Aldrian and Dwi Susanto, 2003; Susilo et al., 2013). Research from Susilo et al. (2013) has shown that during the La Niña years and the ordinary years the average of rainfall in the dry season is three times and two times greater than the amount of rainfall in El Niño years, respectively. Due to the low amount of rainfall in the El Niño years, the fires in those years have a higher frequency compared to non El Niño years (see Figure 2.1). The low amount of rainfall during the El Niño years has a strong influence on the occurrence of fires in Central Kalimantan. However, recent findings showed that peatland fires are now becoming a regular feature of the dry season (Page and Hooijer, 2016), which means the fire activities also highly likely to be found in the non El Niño years (Gaveau et al., 2014). Due to this evidence, other supporting factors should be considered in relation to the peatland fire incidents.

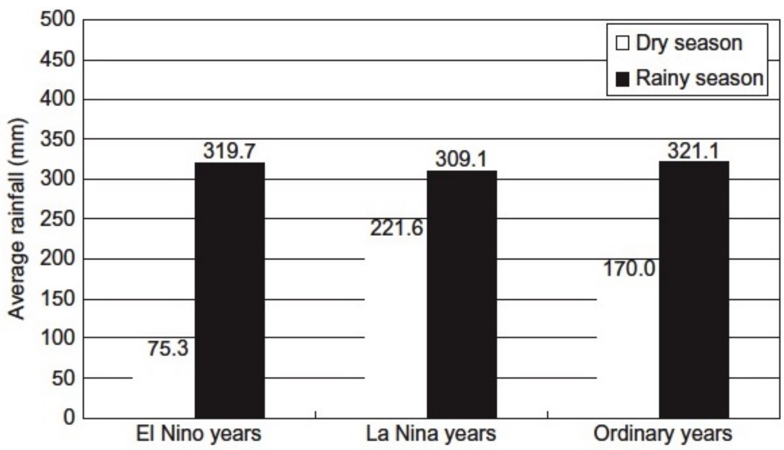


Figure 2.1: Comparison between hotspot in El-Niño, La-Niña, and ordinary years. Sourced from: (Susilo et al., 2013)

One of the critical factors in determining fires occurrence in peatland areas is the ground water level (GWL). Usup et al. (2004) found that when the GWL is very low, the peat and organic materials on the surface became extremely dry and combustible. The study from Putra et al. (2011) found that the peak of the fire period is started when the GWL reaches -40mm from the surface. Therefore, the value of -40mm from the surface is then used as the threshold of the critical GWL in Indonesia. The latest research from Hayasaka et al. (2016) has found a strong linear relationship between the GWL and their calculations of the peat fire index (PFI). PFI is used to estimate the depth of the combustible peat layer. In the years with low GWL, the peat fire indices become high (see Figure. 2.2). The variability of GWL on peatland area is strongly influenced by the amount of rainfall. Even though the nature of peatland is to retain water, the GWL in peatland areas gradually decreases during the dry season. An interesting finding on the relationship between the changes of rainfall with the pattern of GWL and fire occurrence was discovered by Putra et al. (2011).

They found that there is one-month time-lag between the precipitation changes and GWL. The lowest recorded GWL occurred one month after the lowest level of precipitation. The fire occurrence also reached a peak at the time when the GWL showed its lowest value. Despite this one-month time lag, the GWL does gradually decrease while the amount of rainfall decreases. The GWL slowly returns toward the peatland surface after the continuous heavy rainfall happens at the end of dry season.

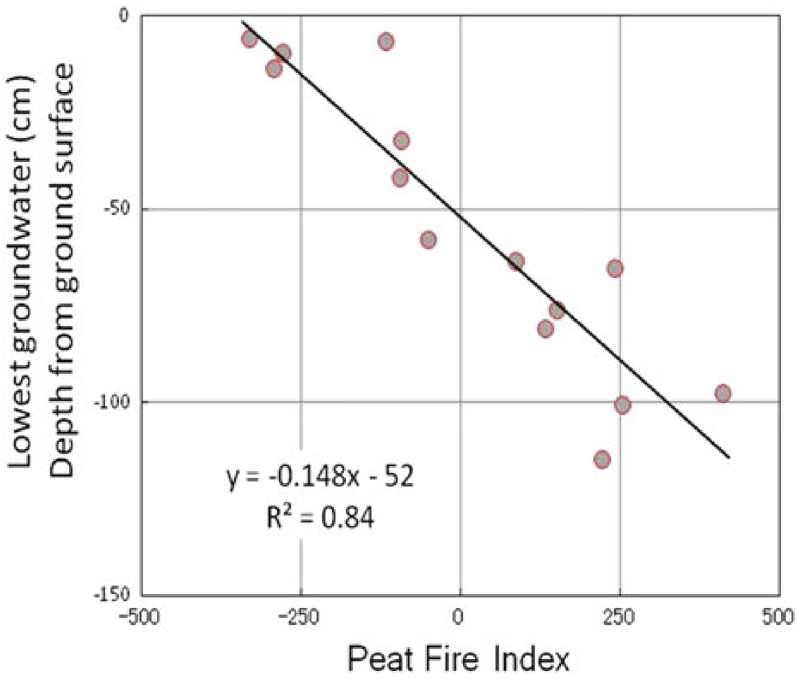


Figure 2.2: The linear regression of GWL and the peat fire index. Sourced from: (Hayasaka et al., 2016)

c. *Physical Condition*

Even though most of the evidence shows that climatic conditions have exacerbated the occurrence of fire over the past two and three decades, Page et al. (2009) established that the changes in the land cover and land use also should be considered as critical factors that increase fire frequency.

The changes in land use happen when peatland areas are converted to settlement, agriculture, plantation, and mining. This has caused a severe degradation of peatland (Adinugroho, 2005). One of the significant causes of peatland conversion is transmigrants' housing and agriculture that began in early 1996 when the Indonesian government initiated a food crop plantation project to support rice production in the country (Page et al., 2009), known as the Mega Rice Project (MRP). Notohadiprawiro (1998) reported that 1 million hectares of peatland were dug out and several thousands of kilometers of drainage channel was built to drain the excess water from the new agricultural area and also as the irrigation channel.

However, this project was eventually discontinued, and almost 1 million hectares of peatland area was left abandoned and degraded. The implication of the discontinuity of this project has been well studied. Adinugroho (2005) investigated the impact of the construction of a drainage channel in the ex-MRP area. They found that the development of the canals has caused the peat to dry out excessively during the dry season and to catch fire quickly. Not long after the conversion project, more than half of this area extensively burned during the 1997 fire season and most of the fires in Central Kalimantan were found in this area (Page et al., 2002; Langner and Siegert, 2006). A strong association between fire risk and a drainage channel was also confirmed by (Wösten et al., 2008). After employing hydrological modelling to investigate this inter-relationship, they estimated that an area with more drainage channels has a lower GWL compared to an area that has an unmodified hydrology. This low water level makes the surface peat sufficiently dry for fire ignition (Usup et al., 2004).

The conversion in the peatland area has changed the structure of land cover from forested to non-forested. This non-forested area is one that is most prone to fires and based on the remote sensing analysis, fires most often happen in the non-forested area. For example, during the 1997-1998 fires in East Kalimantan, 59% of fires occurred in the logged forest and only 5.9% in the undisturbed forest (Siegert et al., 2002; Page et al., 2009). The recent fires of 2006 and 2015 also mostly occurred within the non-forest areas (Harrison et al., 2009; Field et al., 2016). In addition to that, fires also seemed to be found in the areas that had experienced fires in the past. It can be assumed that previously burned areas are more prone to the repeated fires in the future (Langner and Siegert, 2009)

Based on the explanation above, it is clearly seen that forest fires in Indonesia are influenced by three major factors: a suitable fuel environment that makes the land easily catch fire; the human activities that trigger fire ignition; and the weather conditions that support fire ignition and sometimes encourage the fires to spread. Figure 2.3 shows a conceptual model of contributing factors for peatland fires in Indonesia. This conceptual model shows three categories of contributing factors and the relationship of each factor in encouraging the fires.

2.1.2 The Impacts of Peatland Fires

Peatland fires are considered as a potential threat to the sustainable development of Indonesia due to the direct effects on ecosystems and biodiversity and the contribution to the carbon emission (Siegert et al., 2002; Hoscilo et al., 2011; Turetsky et al., 2015). Peatland fires also have significant indirect effects on the people lives and the

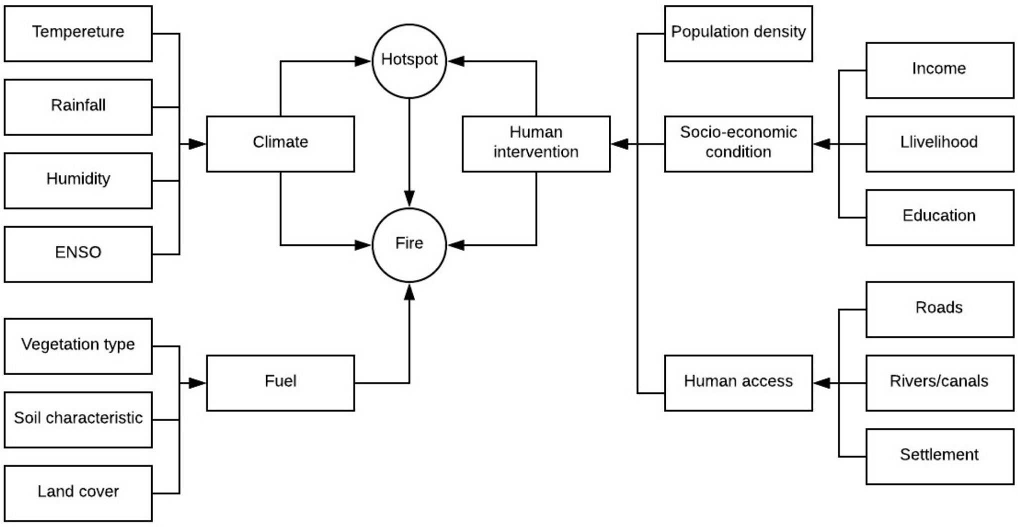


Figure 2.3: Conceptual model of factors contributing in peatland fire in Indonesia

stability of the economy in a country (Harrison et al., 2009; Aditama, 2000). In this subsection, we discuss the cost caused by peatland fires from human life, biodiversity, and economic perspectives.

People

The impact of wildfires on human lives can be direct and indirect. The direct impact mostly occurs in the case of firefighters or residents near a forest. Australia's Ash Wednesday Fires in 1983 resulted in 75 human deaths Goldammer (1998). More than 20 years later, in February 2009, the Black Saturday fires in Victoria, Australia caused the death of lives of 173 people because there was no time to escape from the fires (Whittaker et al., 2013). In the 2003 Southern California fires, 22 residents died because they were trapped by the fires(Mutch, 2007).

In Indonesia, fires mostly happen far away from settlements. Thus, there are no human deaths caused by the direct impact. However, the impact of wildfires on human lives is mostly indirect through the exposure to air pollutants during or after the fires (Cochrane, 2010). Peatland fires are known as a substantial contributor to smoke haze pollution (Heil and Goldammer, 2001; Usup et al., 2004). The smoke haze released from the peatlands has caused poor air quality across densely populated regions in Indonesia. During the peak period of fires in 1997, the total suspended pollutant value in Central Kalimantan and Jambi was more than 15 times the dangerous level (Aditama, 2000). Poor air pollution also happened in Palangka Raya, the capital city of Central Kalimantan in 2006. For more than 81% of the days in September-November, the air quality was rated as unhealthy/very unhealthy and dangerous to human health (Harrison et al., 2009). The recent fires in 2005 also contributed to the worst air quality in 20 years in Central Kalimantan, where the pollutant index was two orders of magnitude higher than the WHO air quality guideline (Organization et al., 2006; Stockwell et al., 2016). The smoke haze has affected not only the Indonesian region but also its neigbouring countries. The transboundary haze from Indonesia has affected the air quality in Singapore and Malaysia (Shahwahid H.O and Othman, 1999; Forsyth, 2014; Othman et al., 2014). In particular, the haze during the 1997 fire season saw the Pollutant Standards Index level increase to an unhealthy and hazardous condition for most of the days in September-November in Singapore (Hon, 1999)

The particles in the smoke haze resulting from peatland fires are dangerous to human health. The chemical pollutants can cause serious health problems such as acute respiratory infection (ARI), bronchial asthma, skin disease, and eye irritation (Aditama, 2000). Based on the data collected from the governmental health service, there was an increase in the number of patients who suffered ARI and bronchial asthma during the period of fire in 1997/1998 (Aditama, 2000; Kunii et al., 2002). The impact of smoke haze on human health can last for several months after the smoke has disappeared, especially for women (Frankenberg et al., 2005).

The chemical pollutants also cost human lives during the smoke haze period. The report from the Indonesian Minister of Health recorded 29 deaths in Central Kalimantan during the fire season in 1997 (Limin et al., 2007). The mortality rate was also increasing two to four times in Jambi during that fire season, mainly due to respiratory problems (Aditama, 2000).

Biodiversity

In sustainable development, fires are considered a potential threat because of their direct impact on biodiversity. The immediate damage caused by the fires is the destruction of the natural vegetation and the wildlife in the forest. Deforestation is one of the destructive impacts of fires that threaten the existence of biodiversity in the forest. The 1997-1998 fires in Indonesia have been blamed as the largest contributor to the forest loss and deforestation in Borneo (Fuller et al., 2004). In 2000, the annual deforestation rate in Borneo reached 2.0%. Several researchers stated that deforestation is expected to continue in this island (Margono et al., 2014). The impact of deforestation in the Borneo forest is massive because of the high biodiversity and large extent of tropical peatland in this region(Miettinen et al., 2011).

Deforestation has threatened the existence of some indigenous animals such as some species of birds, orangutan, gibbon and leaf monkey. These animals have lost their habitat and die from lack of food and water (Harrison et al., 2009). It was reported in Kinnaird and O'Brien (1998) that after the fires, the number of frugivorous birds decreased by 13%. After the fire season, most of the trees failed to produce fruit. That caused difficulties in finding food for these birds. Rijksen and Meijaard (1999) reported that one-third of the remaining orangutan were estimated dead after the severe fires in 1997/1998.

Economic

The cost for suppressing fires, the economic loss in agriculture and forest resources, and also the indirect economic loss caused by the haze has impacted the Indonesian economy. The Asian Development Bank (ADB) together with Bappenas (the Indonesian National Planning Agency) estimated that the economic cost of 1997-1998 fires reached more than US$ 9 billion (Applegate et al., 2002). The major economic loss was from the loss of crops and timber in the natural forest and plantation. Tourism as one of the contributors to the Indonesian economic sector had also significantly reduced over the period of the fires. The smoke haze has caused the closing down of the airports and delays in the airline schedules. It was estimated that during the 1997-1998 fires, the damage from fire-related and haze-related damage might have been as high as 5 % of Indonesia's gross national product (GNP) in 1997- 1998 (Cochrane, 2010).

Almost 20 years later, severe fires happened again in Indonesia. The World Bank estimated that the 2015 fires cost Indonesia the equivalent of at least US$ 16.1 billion amounting to 1.9 % of Indonesian gross domestic product (GDP) in that year (Bank, 2016). The World Bank also estimated that the losses in agriculture and forestry are the most significant contributors to the economic losses caused by fires (Bank, 2016). Similar to the impact of the smoke haze of the 1997-1998 fires, the tourism sector also lost a significant revenue. The poor visibility again caused the closing down of the airports and interrupted airline schedules for weeks in a few provinces such as Central Kalimantan and Riau. The smoke haze also interrupted the transportation sector, through delayed cargo shipping, and thus contributed to slower growth in trade services, which suffered losses of more than US$ 1 billion (Bank, 2016).

The economic cost of wildfire is estimated to be even greater if carbon emissions are included. As a result of the peatland and forest fires in Indonesia in 1997-1998, between 0.81GT and 2.57GT of carbon were released to atmosphere (Page et al., 2002) which is equivalent to US$ 60 billion and US$ 190 billion. The recent fires in 2015 also recorded a large amount of carbon that was released into the atmosphere. Within two months, September and October 2015, 0.8 GT of carbon was released by the fires (Huijnen et al., 2016). The Global Fire Emissions Database (GFED) provides a higher estimation of carbon emissions during the 2015 fires. From July to November 2015, the 2015 Indonesian fire contributed roughly 1.75GT of carbon to the atmosphere (Bank, 2016).

2.2 Forest Fire Management Systems

This section aims to address **R.Q 1** as regards the lesson learned from the current approaches implemented by the Indonesian government to tackle the peatland fire incidents. In this section, the current approaches of fire prediction systems in Indonesia and other parts of the world are presented. The limitation of these approaches is presented. The best approach that should be implemented to tackle the peatland fire problems is proposed.

Enhancing the forest fire management system is a priority taken by the Indonesian government to minimise the impact of this fire. Forest fire management can be defined as activities that get the right amount of fires in the right place and at the right time (Martell, 2007). The activities are conducted not only by the fire authorities, but also other stakeholders involved in forest management such as industry, non-profit organisations, and the local community. All of the stakeholders should be responsible for dealing with the forest fires and its impact on people and ecosystems. The fire authorities should be able to implement a number of activities: provide information when and where the fires might occur and provide the prevention mechanism to reduce the impact of the fires; attempt to detect fires while they are small and suppress them before become large; deploy the firefighters and other suppression resources close to the fire areas; manage the use of fires for specific purposes and limited scale; provide the mitigation mechanism when the forest fires do occur (Adinugroho, 2005; Martell, 2007).

Since wildfire is not a new phenomena and the impact is also destructive in different aspects of human life, many countries have developed different methods to suppress fire, extinguish fire while it is small and also prevent fire before it happens (Cochrane, 2010). However, the priority of forest fire management systems is directed more to the suppression of fires (Adinugroho, 2005; North et al., 2015; Mateus and Fernandes, 2014; Herawati and Santoso, 2011). Unfortunately, the endless cycle of fires has sometimes burdened the economic stability of a nation because it has to provide a portion of its budget to suppress the fires (North et al., 2015; Thompson et al., 2015; Mateus and Fernandes, 2014). During the fire season of 2000, the United States spent more than US$ 1.4 billion in fire suppression costs (Tacconi, 2003). The fire suppression cost in Portugal also increased every fire season. In 2010 the fire suppression cost around 94% of the total fire management funding (Mateus and Fernandes, 2014). In Indonesia, the policy and activities related to fires emphasize more on the fire suppression rather than fire prevention. For example, the government institution will only take action when fires have already occurred, which is sometimes too late because the fires have become extensive and difficult to extinguish (Adinugroho, 2005). This leads to massive funding and resources to extinguish the fires and minimise the impact of the fires (Herawati and Santoso, 2011).

Many studies agreed that fire suppression is not only constantly costly and might burden the nation financially (Butry et al., 2001; North et al., 2015; Purnomo et al., 2017), it also sometimes fails to prevent fires from escaping containment lines and extensively burning the forest under extreme weather conditions (North et al., 2015). Thus, they suggested the fire authorities should consider shifting from fire suppression to fire prevention (Purnomo et al., 2017; North et al., 2015; Mateus and Fernandes, 2014). The next subsection discusses the benefit of having a fire prevention system and the implementation of a fire prevention system in different countries including Indonesia.

2.2.1 Fire Prevention Systems in Indonesia

The focus of forest fire prevention systems is how to predict the occurrence of forest fires to minimise the losses caused by the fires. It is an early activity in the process of fire control and should be carried out continuously. Fire prevention is one the most efficient ways of reducing the damage and loss arising from fire, without having to use expensive equipment. The activities in preventing fires consist of efforts to prevent or reduce the risk of fires from escalating to the forest area, to prevent the fires from occurring inside the forest, and prevent small fires becoming wildfires (Adinugroho, 2005).

Fires are not a new phenomenon in Indonesia, especially for fires occurring in the peatland area. Even though the Indonesian government prioritises more on the suppression of the fires (Adinugroho, 2005), there are some actions that have been taken in preventing the fire occurrence, using conventional techniques and also implementing the modern technology. Conventionally, the fire prevention techniques are conducted through the direct observation in the field of locations that are prone to fires. Fire lookout towers are built to support the direct observation (Chandrasekharan, 1998; Wibowo et al., 1996), and the information of the danger of fire breaking out is issued to the villagers through the use of drums or "kentongan" (Adinugroho, 2005).

Through the modern technology, it is possible to develop a fire information system to support prevention of the occurrence of fires. The fire information system can take into account all the factors which affect the incidence of forest fire such as fuel conditions, climate condition, and the fire behaviour. Some fire information systems have been implemented in Indonesia to provide a warning of the possibility of fires such as an early warning system, fire danger rating, and hotspot monitoring.

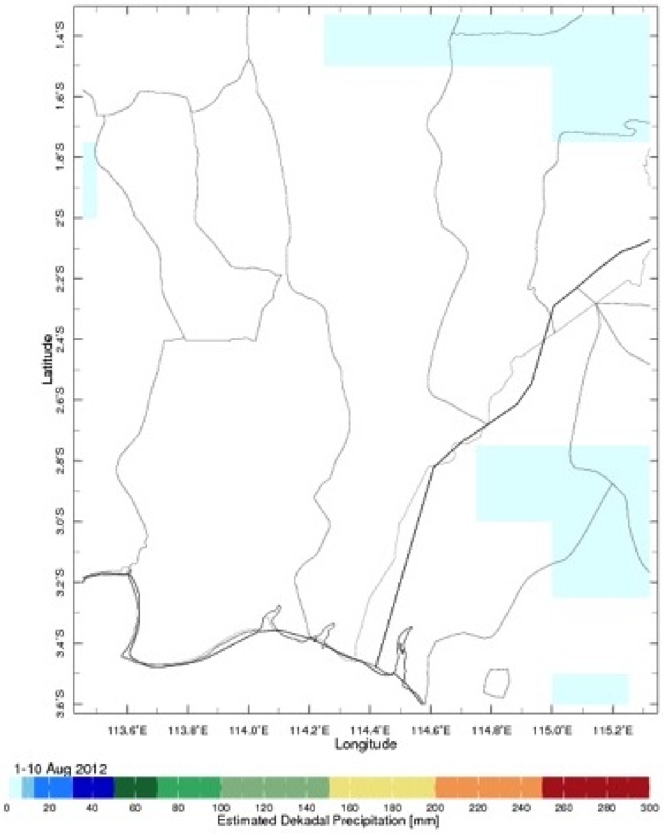
Early Warning System

International Research Institute for Climate and Society (IRI)-Columbia through a partnership with Bogor Agriculture Institute (IPB) proposed tools to monitor rainfall anomalies and forecast the fire occurrence (Ceccato et al., 2010). In the rainfall monitoring tool, the periods of rainfall anomalies (when the rainfall is above or below average) is calculated and the estimation of the amount of rainfall in 10 days is presented in the graphic display. As shown in Figure 2.4(a), the estimation of amount rainfall from 1 - 10 August 2012 is presented. During that period, it is estimated that the amount of rainfall is really low, below the threshold that allocated by the fire authorities. Therefore, the purpose of this tool is to assist stakeholders to assess the likelihood of high or low fire activity in an upcoming season based on the estimation past and future rainfall. The researchers involved in the IRI project found that the rainfall anomalies during the dry season were particularly critical in determining the fire activity. The prediction tool for fire activity in Central Kalimantan, Indonesia was also introduced by the IRI project. This tool estimates the likelihood of high or low fire one to two months in advance, by monitoring the Niño 4 index of the sea surface temperature from April-September. The analysis is based on the data from 1998-2006 on fire hotspots, derived from National Oceanic and Atmospheric Administration-Advanced Very High Resolution Radiometer (NOAA-AVHRR) and on the Niño4 index. The graph in Figure 2.4(b) presents a time series of the Niño4 index. This time series shows relationship of fire activities to the index. Based on the historical data, it is clearly seen that when Niño4 index was high, the fire activity increased. The user then may use this too to forecast 1-2 months ahead whether fire activity is likely to be above or below median levels based on the forecast of Niño4 index.

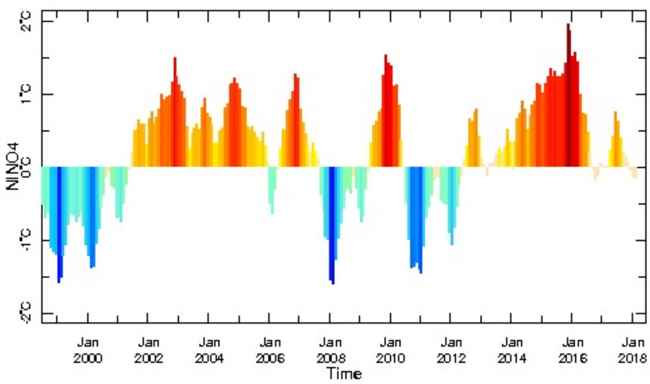
Fire danger rating (FDR)

One of the earliest fire danger rating systems implemented in Indonesia was proposed by the IFFM-project (Deeming, 1995). This FDR applied the Keech-Bryam Drought Index (KBDI) to estimate the drought index. A drought index was obtained through the calculation of KBDI and then interpolated with the East Kalimantan map to generate the fire danger rating maps for East Kalimantan (Hoffmann et al., 1999).

Another FDR that has been implemented in Indonesia is a Fire Danger Rating System (FDRS) which is an adaptation of the Canadian Forest Fire Danger Rating System (CFFDRS). The CFFDRS includes the Canadian Forest Fire Weather Index (FWI) and Canadian Forest Fire Behavior Prediction (FBP) system. The FDRS provides early warning of the potential fire and haze events. The haze or smoke potential indicator was developed using the Drought Code (DC) of the FWI, while the fire ignition potential indicator was developed using the Fire Fuel Moisture Code



(a) Result from the rainfall analysis tool



(b) Result from the fire activity tool

Figure 2.4: Rainfall analysis and fire activity tools proposed by IRI project. Retrieve from: http://iridl.ldeo.columbia.edu/maproom/Fire/Regional/Indonesia/Dekadal_Rainfall.html?region=bb%3A90%3A-12%3A155%3A10%3Abb

(FFMC) of the FWI system. The daily FFMC values were calculated using local rainfall data from the nearest weather station. The FFMC analyses the dead grass moisture content to determine the threshold of the FFMC. Since this system was designed to fit with the forest type in Canada, a few components in the CFFDRS were calibrated to meet with fire environment and fire problems in Southeast Asia (De Groot et al., 2007).

This calibrated FDRS is now implemented and displayed daily on the Meteorological, Climatological and Geophysical Agency (BMKG) website (see Figure 2.5). The information provided by FDRS is used by the agencies as early warning information and used to prepare the fire prevention guidelines.

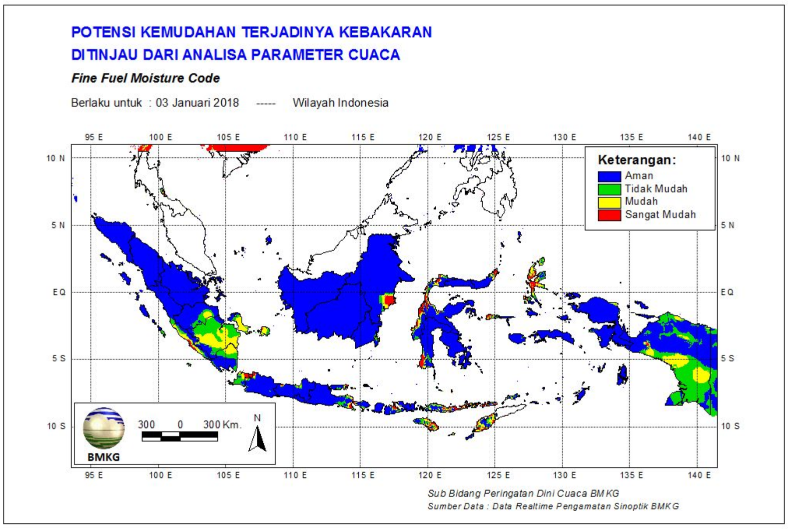


Figure 2.5: Examples of daily FDRS maps in Indonesia. Retrieved from:https://www.bmkg.go.id/cuaca/kebakaran-hutan.bmkg

Hotspot monitoring

Hotspots are high temperature events that sensors on the weather satellites can identify (Langner and Siegert, 2009). Each hotspot could be flagged as containing one or more fires, or other thermal anomalies (such as volcanoes, sun glint, or gas flames on oil platforms). There are a number of satellite and airborne remote sensing devices such as NOAA-AVHRR, Earth and Science Research (ERS), Envisat, and Moderate Resolution Imaging Spectroradiometer (MODIS) Terra/Aqua that can be used to monitor the occurrence of hotspots. In Indonesia, the hotspot data from NOAA-AVHRR and Terra/Aqua from MODIS are mainly used to monitor the hotspot occurrences. Many fire agencies rely on the hotspot information from the satellite as an indicator of fire occurrence. National Institute of Aeronautics and Space (LAPAN), as one of the agencies responsible for managing and distributing the hotspot information, releases a web-based map to show the daily information on hotspot occurrence (see Figure 2.6). Based on this map, the stakeholder can identify the location of hotspots and the level of confidence. Since not all hotspots can be identified as fires, LAPAN also issued a guideline to identify which hotspots were fires.

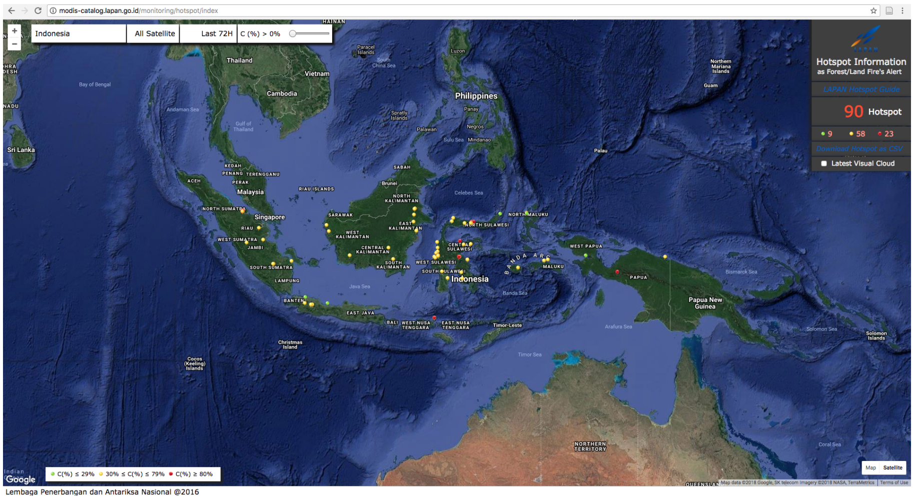


Figure 2.6: Web-based map of daily hotspot information. Retrieved from: https://www.lapan.go.id/

Table 2.1 shows a summary of the three approaches implemented in Indonesia. Most of them only take into account the climatic condition such as rainfall and temperature. Even though the fire danger rating system (Hoffman et al., 1995; De Groot et al., 2007) expanded variables to include the condition of vegetation fuels, the unique characteristic of peatland fires still has not yet been captured. As mentioned in section 2.1.1, the nature of anthropogenic fires is the most influential factor in peatland fires occurrence (Page and Hooijer, 2016). Human activities and their involvement in creating fires should be included as one of the variables in understanding and prevention of the peatland fires.

Table 2.1: Summary of approaches implemented by Indonesian government to manage forest fire incidents

Approaches	**Variables**	**Fire Autorities**
Early warning system	Rainfall, Niño4 index sea-temperature	IRI
Fire danger rating	Rainfall, vegetation fuel	BMKG Indonesia
Hotspot monitoring	Hotspot/satellite imagery fuel	LAPAN, BKSDA

Through examination of the literature, it was identified that none of the officially implemented approaches consider human activities as a contributing factor for peatland fire occurrence in the fire prevention system. Based on the literature review and discussion with the experts, it was determined that the lack of incorporation of the comprehensive factors of peatland fires' characteristic has resulted in the implementation of the fire management system being less successful than desired.

This thesis also explored the fire prevention systems around the world to find the most suitable approaches that could be implemented for peatland fires in Indonesia. The analysis of the approaches around the world is presented in the next section.

2.2.2 Fire prevention systems around the world

Forest fires are a worldwide problem. Many countries experience fires and suffer the impacts. Therefore, the fire prevention systems such as early warning systems or fire danger ratings have been developed in many countries. This literature review looked at fire danger rating systems implemented in three different countries. These systems were also widely adopted in other countries around the world.

Canada

The Canadian Forest Fire Danger Rating System (CFFDRS, (Stocks et al., 1989)) is used across Canada each day during the fire season to help the fire management make decisions regarding the prevention of wildfires. There are two major subsystems in the CFFDRS: the Fire Weather Index (FWI) and Fire Behaviour Prediction (FBR) system. The FWI system relies on the daily weather conditions and relates this information to fuel moisture and fire danger indices for a standard forest type (mainly it is a pine forest type) (Fujioka et al., 2008).

The CFFDRS deals with the prediction of fire occurrence and behaviour from a single source measurement (Tian et al., 2005). The system does not take into account the measurement for spatial variation in the weather elements. External models or systems need to be added to CFFDRS in order to handle this interpolation. This limitation is worthy to emphasize because sometimes it is difficult to obtain an accurate and timely forecasting or estimation of weather conditions in one area. As a whole or in part, CFFDRS calculations depend on the weather conditions.

Despite its limitations, the CFFDRS is one of the most well developed and widely applied schemes. Even though it was designed to fit the condition of boreal and temperate forests and their weather conditions, many countries in other regions have adopted and implemented this system such as New Zealand, Fiji, Mexico, USA (in Alaska), and Southeast Asian countries(Taylor and Alexander, 2006). However, Taylor and Alexander (2006) suggested that when a country is adopting CFFDRS, further research on the characteristics of fuel and fires in that country be undertaken to ensure that the system takes into account the fire behaviour in that country.

United State of America

The United States of America is one of the developed countries that also experiences repeated wildfires. The first National Fire Danger Rating system (NFDRS) was introduced in 1978 with optional revisions added in 1988 (Hardy and Hardy, 2007). This system is used by all Federal and State natural management agencies to measure the fire potential in wildlands. Similar to the CFFDRS, NFDRS is estimated using the measurement of fuel moisture content based on weather conditions such as relative humidity, temperature, and precipitation (Cohen and Deeming, 1985). The American system of fire danger is best implemented in the open, grassy forests or brush types with little or no duff layer (Van Wagner, 1975). This forest condition is different to the tropical forest in Indonesia, especially the peatland area. Therefore, it might not be suitable for Indonesian peatland fires.

Australia

The fire danger rating in Australia is determined by the McArthur model for grasslands (McArthur, 1966) and eucalyptus forest (McArthur, 1967). The forest fire danger rating system is based on the predicted rate of fire spread on dry forest litter and the difficulty of suppressing the fire under certain weather conditions. The rating consists of categories of the Forest Fire Danger Index (FFDI) and the Grassland Fire Danger Index (GFDI). The FFDI relies on the weather conditions. The GFDI is not only based on the weather condition, but also on the proportion of dead grass (Noble et al., 1980). Based on this fire danger rating, the fire warning is issued by the Bureau of Meteorology to the public. Based on this warning, the land management agencies declare the Total Fire Ban policy which does not allow any lighting of fire in the open area (Buxton et al., 2011).

Other forest fire danger rating system

Beside Canada, the USA, and Australia, some other countries also derive their own fire danger ratings. For example, New Zealand, Southeast Asia countries, and the European Union. However, most of them built the fire danger rating based on the systems mentioned above.

2.2.3 Recommendation for Peatland Fire Prediction System in Indonesia

Due to the complexity of the characteristics of peatland fires, it is essential to possess a comprehensive knowledge before developing a prediction system for the escalation of peatland fires. To be able to build a model that shows interaction between human and non-human systems, a flexible and multidisciplinary modelling approach is required (McCann et al., 2006). Multidisciplinary modelling approaches related to the fire management systems have been implemented in different regions. However, Goldammer (1998) argues that fire management systems which include prevention and early warning cannot be generalized due to the different characteristics of fires and their effects. The vegetation zone and ecosystem as well as the cultural, social, and economic factors are the influential factors in the characteristics of fires. Thus, it is important for each region or nation to consider building or developing their own fire management system that can accommodate the characteristics of the wildfire (Taylor and Alexander, 2006).

From the literature review, it was also found that fire monitoring in Indonesia relies on hotspot information from satellite imagery (Roswintiarti et al., 2016). Thus, research in fire modelling approaches is mostly aimed at predicting hotspot occurrence as an indication of forest fires. However, not all hotspots are detected in land and forest will escalate to widespread fires (Vayda, 1999). Therefore, Saharjo (2016) suggested that any fire prediction system should be able to identify hotspots that potentially could turn into forest fire. The prediction system also should be able to deliver accurate information about the location of forest fire once a hotspot is detected.

2.3 Knowledge Elicitation

Knowledge elicitation is the process of acquiring knowledge. This process includes all the activities to obtain knowledge from any relevant sources. Cordingley (1989) mentioned that the knowledge could be obtained from written materials. In this thesis, literature such as articles, letters, memo, reports, or procedural manuals are used. However, (Cooke, 1994) believes that elicitation from human sources is still needed to verify and extend the knowledge obtained from written materials.

Hoffman et al. (1995) proposed four stages of knowledge elicitation. It starts with eliciting the domain knowledge, then generating a first-pass knowledge base, followed by validating and refining the knowledge base. The last stage is instantiating the refined knowledge base in documents or implementable systems. All the activities in the four stages involve information from both written materials and human experts. Thus, in this subsection a further explanation of how to elicit the knowledge from the written materials and human experts is provided, the challenges that might arise in the process of elicitation, and also the justification for choosing the elicitation method for this thesis.

2.3.1 Information Gathering Through Literature

When it is available, the information in the literature is one of the key sources of knowledge that should be considered in the knowledge elicitation. Literature is a useful starter because it reflects a consensual view in the domain (Aussenac-Gilles et al., 2000). Knowledge elicitation from the literature is essential, especially in gathering the basic knowledge of a domain problem. Hoffman (1987) analysed the available documents to generate the basic concepts and definitions in the aerial photography before they conducted further elicitation from the experts. Hickey and Davis (2003) also generated preliminary information on the situations that affected the selection of elicitation techniques from a selection of writing. Even though using literature as the source of knowledge is sometimes indispensable in the knowledge elicitation, the process can be time-consuming (Tang et al., 1994; Hoffman et al., 1995). Thus, automatic knowledge elicitation from literature should be considered as the solution to this challenge.

2.3.1.1 Text Mining

Text mining can be defined as a process of knowledge discovery from a textual database. Generally, it refers to a process of extracting patterns or knowledge from structured documents (such as relational database management systems (RDBMS)), semi-structured text documents (such as eXtensible Markup Language (XML) and JavaScript Object Notation(JSON)), and unstructured text documents (such as word documents, videos, and images) (Tan et al., 1999; Gupta et al., 2009). Text mining can be visualized as consisting of two phases: text refining and knowledge distillation (see Figure 2.7). Text refining is a process to transform the free-forms of text documents into a chosen intermediate form (IF). An IF can be semi-structured such as the conceptual graph representation or structured such as the relational data representation. An IF also can be in a document-based form where each entity represents a document or a concept-based form which represents a concept of interest in a specific domain. Knowledge distillation aims to deduce the hidden patterns or knowledge from the IF.

Text mining covers a large set of related topics and algorithm for analyzing the text documents, including information retrieval, natural language prepossessing, and information extraction from text; supervised or unsupervised learning methods; and text summarization. This thesis implemented the unsupervised learning methods as a technique to discover the hidden pattern from unlabeled data. Topic modeling as one of the unsupervised learning algorithms is used to elicit the knowledge from the text documents.

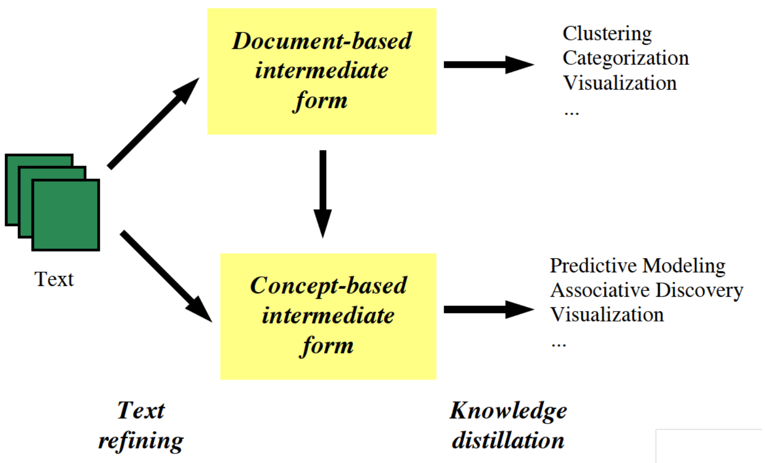


Figure 2.7: Structure of a text mining framework (Tan et al., 1999)

.

2.3.1.2 Topic Modeling

In topic modeling a probabilistic model is used to analyse the pattern of words used in a collection of documents (Blei, 2012). By discovering the pattern, the topics that run through the documents and also the connection between those topics can be discovered. Topic modeling treats each document as a bag of words and identifies words that tend to co-occur in the whole document (Jockers, 2014). By discovering these patterns, we can easily find the topic of interest, then organise and summarise a collection of documents on a large scale without spending a huge amount of time and human effort (Blei and Lafferty, 2009). Topic models have been applied to many kinds of documents such as scientific abstracts (Wallach, 2006; Griffiths and Steyvers, 2004), fiction (Wu et al., 2017), email (McCallum et al., 2005), and newspaper archives (DiMaggio et al., 2013).

A variety of probabilistic topic models have been used to analyse the content of documents and the meaning of words. The simplest topic model is Latent Dirichlet Allocation (LDA) introduced by Blei (2012). The basic idea of LDA is that multiple topics can be revealed in documents. For example, consider the article entitled "Combustion and thermal characteristic of peat fire in tropical peatland in Central Kalimantan, Indonesia". This article is about the characteristic of peatland fire. Manually, we can discover topics in an article by highlighting different words that are used in the article to indicate that the word belongs to specific topics (see Figure 2.8). Words related to ecology conditions, such as *tropical*, *peatland*, *carbon*, and *peat*, are highlighted in yellow; words about climate conditions, such as *El Niño* and *drought* are highlighted in pink; words about human activities, such as *human*, *anthropogenic*, *logging*, *settlements* are highlighted in blue. If we continue to highlight every word in the article (excluding the common terms such as, the, is, are), the words in each topic will increase and the number of words on each topic will vary. This process reveals that one article can contain different topics in different proportions.

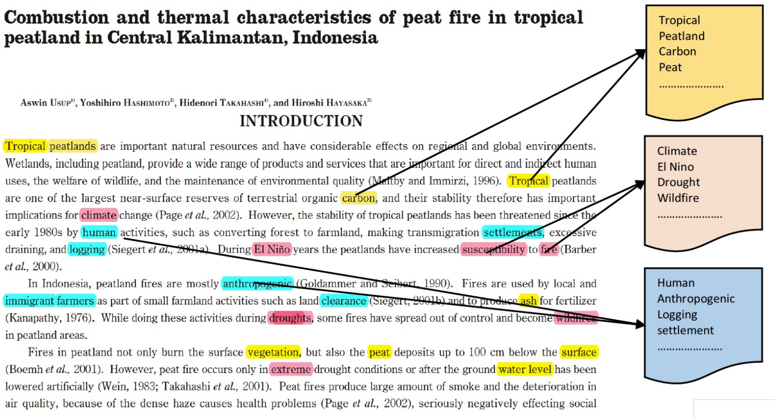


Figure 2.8: Manual process of capturing topics

The manual process of capturing the topics, as explained above, is the intuition that tries to be captured by LDA (Blei, 2012). Each topic has internal consistency, where words of the same topic tend to occur together in the document and do not appear much outside the topic. The words of each document are the observable data, while the topic structure such as the topic itself and how each document exhibits the words are the hidden variable. The interaction between the observable data and hidden topic structure is manifest in the probabilistic generative process associated with LDA.

Figure 2.9 shows the generative process of LDA in the graphical model proposed by Blei (2012): Firstly, LDA defines K topics, with each topic k associated with the distribution ψ_k over the words of topic k. Based on these created topics, observable words w_d is generated by sampling a distribution θ_d over topic K from Dirichlet prior on the per-document topic distribution α. Distribution θ_d can be used to determine the topic assignment for each observable word w_d. Each word $w_{d,n}$ is chosen based on the distribution θ_d. The LDA process can be summarised in three steps (Sun and Yin, 2017):

Step 1: The distribution over the words of topic k is determined by ψ_k.

Step 2: Topic distribution for each document d is determined by θ_d.

Step 3: For each document d, for each word $w_{d,n}$ in d.

- Choose a topic from the distribution over the topic.
- Choose a word from the corresponding distribution over the documents.

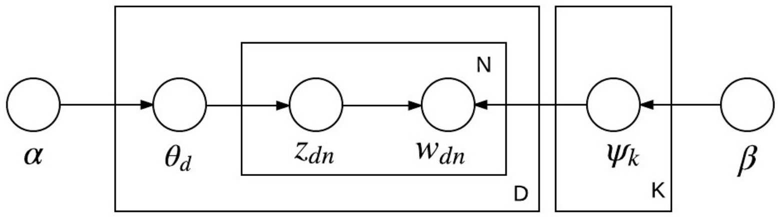


Figure 2.9: Graphical model representation of LDA (Blei, 2012)

.

Posterior distribution is a conditional distribution of the hidden variables given in a document. Computing this distribution is a problem that needs to be solved in order to use LDA. A wide variety of inference algorithms can be considered to compute the posterior distribution for LDA, such as Laplace approximation, variational approximation, and Markov Chain Monte Carlo.

Variational Expectation Maximization (VEM): VEM is the deterministic variational EM method that computes the term distribution of the topic and the topic distribution of documents via expectation maximization. For topic models, the posterior distribution is replaced by a variational distribution.

Gibbs sampling: Gibbs sampling is a special case of Markov Chain Monte Carlo, a simple algorithm that can be used in the high-dimensional models such as LDA (Heinrich, 2005). The posterior distribution using Gibbs sampling is drawn from the formula below (Griffiths and Steyvers, 2004).

$$p(Z_i = K|w, z_{-i}) \propto \frac{n^{(w_i)}_{-i,K} + \beta n^{(d_i)}_{-i,K} + \alpha}{n^{(\cdot)}_{-i,K} + W\beta n^{d_i}_{-i_p} + K\alpha} \tag{2.1}$$

z_{-i} is the vector of current topic membership of all words without the ith word w_i. $n^{(w_i)}_{-i,K}$ is the number of times term j from the document d has been assigned to topic K with the ith word. $n^{(\cdot)}_{-i,K}$ is a count that does not include the current assignment of Z_i. d_i represents the document in the corpus to which word w_i belongs. The index j indicates that w_i is equal to the jth term in vocabulary.

The term distribution of the topic ψ and the topic distribution of documents θ can be estimated by value z from the posterior distribution. These values correspond to the predictive distributions over a new word and a new topic on w and z. Gibbs sampling has algorithm that is easy to implement, requires little memory, and is competitive in speed and performance with existing algorithms (Griffiths and Steyvers, 2004).

Model evaluation and selection

A common problem in topic modeling is to choose the number of topics, if this parameter is not specified in the beginning of modeling process (Blei and Lafferty, 2009). A variety of performance metrics has been introduced based on the goals and available means (Wallach, 2006; Blei, 2012; Zhao et al., 2015). The performance can be measured using data, a secondary task, and human judgement.

Most research using topic modeling measures the performance on data based on the estimation the probability of held-out documents. Wallach (2006) stated that this estimation provides a clear, interpretable metric for evaluating the performance of topic models relative to other topic-based models as well as to other non-topic-based generative models. These metrics are well suited for choosing the number of topics that provides the best language model (Blei and Lafferty, 2009).

A variety of method have been used to measure the metrics' performance in topic modeling research. The methods are perplexity, the harmonic mean method, importance sampling, Chib-style estimation, and left-right evaluation. In this research, two performance metrics, perplexity and harmonic mean are conducted to evaluate the models and choose the best number of topics based on its simplicity and computational efficiency.

- Perplexity. Blei et al. (2003) computed the perplexity of a held-out test set to measure how well a probability model predicts a sample. Perplexity consistently decreases the likelihood of test data and computes the the inverse of the geometric mean of per-word likelihood. A lower perplexity score indicates better generalisation performance. The perplexity equation for a test set of documents is:

$$perplexity(D_{test}) = exp\left\{ - \frac{\sum_{d=1}^{M} log\ p(w_d)}{\sum_{d=1}^{M} N_d} \right\} \tag{2.2}$$

  M is the total number of words in the corpus, w_d and N_d indicate the identities and number of words in the test set, respectively.

- Harmonic mean method. This method has been used in variety of topic modeling research due to its simplicity and relative computational efficiency (Wallach et al., 2009). Griffiths and Steyvers (2004) stated that one way to calculate the likelihood of words, w, specified by the number of topics, K $p(w|K)$is by taking the harmonic mean of set of values of $p(w|z, K)$ when z is sampled from the posterior $p(z|w, K)$. Using the Gibbs sampling, the value of

$p(z|w, K)$ can be computed from the equation:

$$P(w|z) = \left(\frac{\Gamma(V\beta)}{\Gamma(\beta)^V}\right)^K \prod_{k=1}^{K} \frac{\prod_V \Gamma(n_k^{(w)} + \beta)}{\Gamma(n_k^{(\cdot)} + V\beta)} \tag{2.3}$$

Topic coherence

There is no guarantee that all of the topics generated by a topic model will be easy to understand and interpret. For example, the two topics presented below have a different degree of human-interpretability:

Topic 1: *rainfall*, *Niño*, *drought*, *temperature*

Topic 2: *management*, *level*, *geophysics*, *flux*

The first topic is related to a weather condition, while the second topic is less clear and may confuse users when interpreting the meaning of this topic. Topic coherence is used to measure the degree of semantic similarity between high scoring words in the topic (Stevens et al., 2012). Various methodologies have been proposed for measuring the semantic interpretability. An indirect approach based on the word intrusion was proposed by (Chang et al., 2009). This evaluation involved human judgment to identify the intruder words. The automatic methods for estimating topic coherence that have been proven to match well with human judgment of topic quality are the UCI measure (Newman et al., 2010) and the UMass measure (Mimno et al., 2011).

Newman et al. (2010) defines the topic coherence score based on the pointwise mutual information (PMI) of pairs of terms, v_i and v_j, taken from topics (see equation 2.4). The PMI measures the statistical dependence between two words based on their co-occurrence over the corpus. The higher the average pairwise similarity between words in a topic, the more coherent the topic. In Mimno et al. (2011), the UMass metric defines the score of topic coherence based on document co-occurrence (see equation 2.5). Where $D(v_i, v_j)$ counts the number of documents containing words v_i and yv_j and $D(v_j)$ counts the number of documents containing j,

$$score v_i, v_j, \epsilon = log \frac{p(v_i, v_j) + \epsilon}{p(v_i)p(vj)} \tag{2.4}$$

$$score v_i, v_j, \epsilon = log \frac{D(v_i, v_j) + \epsilon}{D(vj)} \tag{2.5}$$

2.3.1.3 Finding similar terms

Similarity between two terms is often represented by the similarity of concepts/meanings associated with these words (Li et al., 2003) or the closeness these two words in the surface level (Zhang and Patrick, 2005). The former is referred to as semantic similarity; the latter is known as lexical similarity. A number of semantic similarity methods have been developed. Li et al. (2003) categorised these methods into two groups: edge counting-based (dictionary/thesaurus-based) and information theory-based (corpus-based). The applications of semantic similarity are mostly for detection or correction of spelling errors, handling ambiguity, text segmentation, or image retrieval.

Lexical similarity only focuses on the similarity of two words at the surface text level and ignores the meaning behind the words. One of the most common metrics for computing similarity between two words is the Jaccard coefficient.

Jaccard similarity coefficient of sets

I examined the similarity between the two set of terms by deploying the Jaccard Similarity Coefficient. The Jaccard similarity addresses the problem of finding textually similar documents in a large corpus (Leskovec et al., 2014), usually used for plagiarism checking, mirror pages checking, or finding articles from the same source. The Jaccard similarity of two data sets A and B is the ratio of the size of the intersection of A and B to the size of their union.

$$SIM(A, B) = \frac{|A \cap B|}{|A \cup B|} \tag{2.6}$$

From the formula above, it can be seen that if A is equal with B the similarity coefficient is 1, meaning that dataset A is precisely the same as dataset B. A similarity coefficient of 0 indicates that there is no similarity between the two datasets. The closer the Jaccard similarity coefficient is to 1, the higher the probability of similarity in the two datasets (Niwattanakul et al., 2013).

2.3.2 Knowledge Acquisition in Human Experts

A significant amount of research agrees that eliciting knowledge from documents is a good starting point to understand the domain of study (Aussenac-Gilles et al., 2000; Hoffman, 1987; Hickey and Davis, 2003); other research suggests that human experts possess knowledge and experience that do not appear in the documents (Meyer and Booker, 2001; Hoffman and Lintern, 2006).

Hoffman et al. (1995) defined an expert as someone who has skills or knowledge in a particular topic. This knowledge could be developed from training, research, or even extensive experience in a specific domain (Burgman et al., 2011). The expert also should be able to provide answers or information that can be obtained from other resources (Meyer and Booker, 2001). When involving experts in the study, some challenges might arise that need to be addressed when choosing the right methods to elicit the knowledge, Bell and Hardiman (1989) have mentioned some of the challenges:

- experts' time is precious
- sometimes encouragement is needed to keep the experts talking or providing information
- experts can be bored and impatient during the elicitation process
- experts can have difficulty in articulating the knowledge that they have.

The method used to elicit knowledge from experts are varied and diverse. Based on Hoffman et al. (1995) the method could be categorised into three categories:

- analysis of familiar tasks. This method is usually conducted to investigate the task performed by the expert to solve a problem.
- various types of interview. This method is conducted to gather knowledge based on what the expert said. It can be done through individual or group interviews.
- contrived techniques. This method is used to analyse what the experts do when they are constrained in some new way.

The choice of which method to be used in eliciting the knowledge from experts is depends on the context of the knowledge elicitation project, such as the type of information that needs to be elicited, the experts involved in the process, and also the expertise of the experts. Research which has made comparisons between the elicitation techniques has found that the contrived techniques are more beneficial and performed very well compared to the interviews (Burton et al., 1990; Hoffman et al., 1995). However, the recent finding from Dieste and Juristo (2011) showed the opposite conclusion: the interview has proven a better technique compared to contrived techniques in the majority of cases. They also mentioned that techniques that are more beneficial compared to others depend on the domain being modelled, since the overview of the domain knowledge sometimes cannot be captured in the contrived techniques, Hoffman et al. (1995) suggested that this technique should be used in conjunction with other techniques such as document analysis or interviews. Thus much research which makes use of knowledge elicitation tends to combine multiple techniques. A typical example may be to use document analysis or unstructured interviews to get a broad understanding and build an initial basic model (Hoffman et al., 1995; Hickey and Davis, 2003), with some structured interviews or tasks in contrived techniques, to yield further knowledge about a domain problem (Vayda, 1999; Hickey and Davis, 2003; Dennis et al., 2005). In this thesis, at first the document analysis is used to elicit the domain knowledge and created an initial basic model. The contrived techniques, delivered via survey and group discussion is conducted after document analysis to finalise the structure of model, quantify the model, and evaluate the performance of the model.

Survey

Asking questions is the obvious way to gather information from people. The questions can be delivered in various ways: by conducting face-to-face or telephone interviews, by having questionnaires filled, or by using specific tools (Boynton and Greenhalgh, 2004; Stieger and Reips, 2010). In general, the interview method is preferable and frequently employed in the knowledge elicitation (Knol et al., 2010; Dieste and Juristo, 2011). It has proven to be more efficient (Frey and Fontana, 1991) and can yield some information about the domain concepts and reasoning (Knol et al., 2010). However, when multiple experts and numerous questions are involved in the elicitation process, conducting interviews can be time-consuming and laborious (Hoffman et al., 1995). Another technique used in asking the questions of multiple experts is through questionnaires: a researcher can ask numerous questions to many people in a short period (Neuman, 2002). A survey is also widely used to gather peoples' perspectives in many domains, such as in health and medicine, (Eysenbach and Wyatt, 2002; Bennett et al., 2007; Winstock et al., 2015), education (Chretien et al., 2015), information technology (Kypri et al., 2004), and religious studies (Hill et al., 2007).

The result from a questionnaire can be collected using a mail-out format or web-based survey. In the mail-out format, the printed questionnaire is set through the mail and submitted also by mail (Rea and Parker, 2014). In the web-based survey, the question is available online and the participants can access the survey from anywhere (Stieger and Reips, 2010). In both of these methods, the participants are asked to complete the question on their own and submit the respond within the agreed time. Both mailed questionnaires and web-based surveys have a low response rate compared to telephone or face-to-face interviews (Neuman, 2002). However, a benefit of mail or web-based questionnaires is anonymity; there is no personal contact between the respondents and the researcher. This will help to avoid bias in interpreting the respondents' answers (Neuman, 2002; Rea and Parker, 2014).

Focus group

A focus group is a special kind of interview situation (Neuman, 2002). The purpose of conducting a focus group is to gather a better understanding of how people think of an issue. A focus group is less threatening to many participants, because for some individuals disclosing their opinion is not easy. Having a comfortable and permissive environment is helpful for participants to discuss perceptions, opinions, and thoughts (Krueger, 2014). Since achieving consensus is not the objective of focus group, the moderator or interviewer should be ensure that the permissive and nurturing environment encourages different perceptions and points of view. Exploring maximally the various perspectives held by participants is more important rather than forcing the participants to vote, plan, or reach concensus.

(Krueger, 2014) mentioned five characteristics of a focus group:

1. It is a small group. Typically composed of 5 to 8 people, but can range from as few as 4 people to as many as 12 people.

2. A person involved in a focus group should possess certain characteristics. Participants in a focus group usually have common characteristics that relate to the topics in the focus group. Jourard (1964) has found that individuals decide to reveal their thinking and feeling to people who have something in common rather than people who differ in many ways with them.

3. Qualitative data is provided. The goal of focus group is to find the range of opinions of people involved in the discussion. Therefore, the data collected from a focus group are solicited through open-ended questions.

4. Focus groups have a focused discussion. Open-ended questions are usually used by the moderator in the focus group. These questions should be carefully predetermined and sequenced so discussion remains within the context.

5. The purpose of the discussion is to help understand the topic of interest. It can be in the form of collecting the general background of the topic, identifying the potential problems, or providing assistance with the interpretation of the quantitative result (Stewart and Shamdasani, 2014).

2.3.3 Bias in Expert Elicitation

The role of experts in knowledge elicitation is not to make judgments but to provide clear information about the consequences and probabilities in a problem domain that allows the decision-maker to make a better decision (Pollino and Henderson, 2010). However, the expert opinions are still prone to cognitive and knowledge biases (McBride and Burgman, 2012). Baddeley et al. (2004) and McBride and Burgman (2012) categorised the main sources of the expert bias as motivational bias and cognitive bias. Motivational bias reflects the interest and circumstances of the experts. It can arise from the context, personal belief, or the advantages or risks the expert might gain related to the outcome of the elicitation. However, cognitive bias emerges from the incorrect processing of the information. Cognitive bias usually arises when a human makes a judgement based on common sense or rules of thumb. This judgment is sometimes derived from experience or is known as a heuristic. Heuristics are mostly used to make quick decisions in uncertain situations. However, the problem that might arise when employing heuristics is that the expert is often overconfident about their knowledge. They might overestimate or underestimate the accuracy of their knowledge (Kuhnert et al., 2010). Bias also can occur when the expert only provides judgements based on recent information and the more complex and attractive events, and does not consider the past events or the more frequent events (Baddeley et al., 2004).

Linguistic uncertainty is also one of the cognitive biases that arises because words have imprecise or different meanings. Experts might misunderstand the questions or apply different interpretations to the same question (Kuhnert et al., 2010). Thus, developing the right questions that are asked in the right way, is the key for a survey to succeed in eliciting expert knowledge (Gable, 1994), because good questions will deliver valid and reliable measurements. The result from a survey can be threatened by many factors, such as biased questionnaire design and wording, faulty questionnaire design, and misunderstanding (Harris and Brown, 2010). Neuman (2002) suggested a few techniques to avoid confusion and keep the respondents' perspective in mind in developing the right questions for eliciting the expert's knowledge.

The questions in the survey should:

- be clear and unambiguous. Neuman (2002) mentioned that sometimes a researcher makes implicit assumptions without thinking of the respondents. This might happen because the researchers become so deeply involved in the topic, that the perspective of the questions become really clear for them by not necessarily clear to the respondents (Babbie, 1990).
- use jargon and technical terms widely known by the respondents (Neuman, 2002), expecially when having multiple experts with different backgrounds.
- be relevant to the respondent, related either to their educational background or life experience.

- not be double-barrelled. A double-barreled question is question with two opinions or objects that are joined together. Basically it asks respondents to answer two questions with one answer (Bradburn et al., 2004).
- not be phrased in the negative. Cassels and Johnstone (1984) found that questions with negative forms required more working memory capacity, especially for understanding the questions prior to delivering correct answers.

Having multiple experts with diversity of knowledge and expertise is suggested to avoid individual bias (Pollino and Henderson, 2010). However, Baddeley et al. (2004) said that when information is collected through a group discussion, the group interaction might cause another bias to arise. The mistakes and misjudgment could be communicated between the experts through the group interaction.

2.4 Causal Model

As explained in Section 2.2.3, it is important to understand the behaviour and characteristics of peatland fires before finding solutions to tackle the problems of fires. Identifying the causal relations between the contributing factors in peatland fires can be done to understand this environmental problem. A causal model is used to present the causal relations and explanations between the variables of interest (Russo et al., 2011).

Causal links via a graph or diagram have been widely known in causal analysis (Greenland et al., 1999). A causal graph can consist of a set of structural-equation models or/and a graphical model. In recent years, the graphical model of a causal relationship between variables of interest was also known as a causal diagram implemented as a directed acyclic graph (DAG). One of the applications of causal models using DAGs is found in Bayesian Networks (BNs). In this thesis the understanding of peatland fire characteristics and behaviour is presented in probabilistic graphical model using BNs. Therefore, an explanation on BNs is presented in the next subsection.

2.4.1 Bayesian Networks (BNs)

BNs are probabilistic graphical models that are useful to infer the causal relationship or interactions among a set of random variables and use directed acyclic graphs (DAGs) to represent the causal dependencies (Pearl 1988). BNs are mostly used for reasoning under uncertainty (Korb and Nicholson, 2011) and often used for modeling when explanation of the relationships between the variables are not easily expressed using mathematical notation (Pearl, 2000). BNs are composed of three elements: a set of nodes representing variables (discrete or continuous), a set of links known as arcs representing the direct connections between the variables, and a set of conditional probability tables (CPTs) specifying the belief of the relationship on each node.

In the structure of BNs, a node is a parent of a child if there is a direct arc from the parent to the child. A root node is a node without parents that represents the cause; a leaf node is node without children that represents the final effects. A node which is both non-leaf and non-root is called an intermediate node (Korb and Nicholson, 2011). The link or arc between the nodes represents the causal dependencies based on the process understanding statistical or other types of associations (Pollino et al., 2007).

BNs apply Bayes' Theorem (also known as Bayes' rule or Bayes' law). In Bayes' theorem, a prior (unconditional) probability represents the likelihood that an input node will be in a particular state, P(A) and P(B); the conditional probability calculates the likelihood of the state in a node affected by other nodes, P(B|A); and the posterior probability is the likelihood that a node will be in a particular state, P(A|B), given the input nodes, the conditional probabilities and the rules governing how the probabilities combine. BNs use this theorem to update or revise the beliefs of the probabilities of system states taking certain values, in light of new evidence (Pollino and Henderson, 2010).

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{2.7}$$

BNs emerged from research into artificial intelligence, where they were originally developed as a formal means of analysing decision strategies under uncertain conditions (Varis, 1997). BNs are particularly useful for diverse problems of varying size and complexity, where uncertainties are inherent in the system. In environmental issues, uncertainty in the environmental knowledge becomes a central concern during the process of decision making (Eden, 1998). This uncertainty or the lack of knowledge or information prevents the user from providing a precise measurement of the system or observing the behaviour of the system (Finkel, 1996). BNs are widely used for the analysis of data and expert knowledge in a domain that is full of uncertainty, due to the ability of this method for dealing with uncertainty and complex problems in the environmental domains (Uusitalo, 2007; Pollino and Henderson, 2010). One way to reduce the uncertainty is through adaptive management; this allows the management actions to be adjusted over time. BNs provide a framework for iterative updating as more knowledge becomes available (Pollino and Henderson, 2010). For example, the iterative process happens during calibrating and updating the structure of BNs or specifying the CPTs (Korb and Nicholson, 2011; Marcot et al., 2006). A number of researchers have proven that BNs work best with small and incomplete data sets (Korb and Nicholson, 2011; Uusitalo, 2007; Pollino and Henderson, 2010; Marcot et al., 2006). There are no minimum sample sizes required to perform analysis using BNs. BNs have the flexibility of working with both data-poor and data-rich environments. Even with a small sample size, a BN still can show good prediction (Kontkanen et al., 1997). However, Marcot et al. (2006) warned of the possibility of over-fitting the model to data with a small sample size. Another uncertainty issue in the ecological and environmental domain is missing values. Therefore, the implementation of BNs is a good choice because they can incorporate the missing data through the application of Bayes' theorem (Pollino and Henderson, 2010).

Ecology and environmental domains are complex areas where not only physical or biological knowledge are needed, but also knowledge of interaction with humans. To be able to build a model showing interaction between humans and a non-human system, a flexible and multidisciplinary modelling approach is required (McCann et al., 2006). This approach should be able to assemble the diverse information into a coherent and systematic environment (Pollino and Henderson, 2010). Therefore, BNs are one of the modelling approaches that readily integrate information from a range of disciplines (Pollino and Henderson, 2010), combining the knowledge from experts and data (Marcot et al., 2001; Pollino et al., 2007) and incorporating both quantitative and qualitative evidence across a range of scales(Pollino and Henderson, 2010).

2.4.1.1 The way to construct BNs

There are three steps in developing any BN (Druzdel and Van Der Gaag, 2000):

1. Identification of variables to be included in the BN
2. Identification and representation of the relationships between variables in the network structure
3. Parameterisation of the network

In the theory, these tasks should have been done sequentially (Nicholson et al., 2001). The structure of a BN is only created once the variables are defined. The parameterisation of the nodes should be done once the structure is completed.

However, in the practice of knowledge engineering processing, iteration could occur to ensure the model is considered acceptable, either based on the experts' knowledge or data evaluation.

Learning BNs using historical data

A reasonably sized dataset is required in order to learn BNs using historical data. The size also depends on the complexity of the network (Guo et al., 2017). Without a sufficient amount of data, it might be difficult to produce a BN that accurately shows a causal relationship of a problem.

Methods for learning BNs from data fall into two categories: constraint-based and score-based. Constraint-based algorithms are based on the seminal work of Pearl on causal graphical models and the Inductive Causation algorithm (Verma and Pearl, 1991). This algorithm provides a framework for learning the directed acyclic graph (DAG) of a BN using conditional independence tests under the assumption that graphical separation and probabilistic independence imply each other. The score-based approach is a scoring function that evaluates how well the DAG explains the data and is a way to find the best DAG that optimises the scoring function. Over the last decade, learning BNs from data became more popular, especially with increasing volumes of data available. Various fields such as bio-medical, internet, and e-business are examples of applications that utilise data availability to learn BNs.

Learning BNs using expert knowledge

Learning BNs from data is a preferable method now in various fields, due to being less time-consuming compared to eliciting knowledge from experts. However, in many fields and circumstances, limited or insufficient data is still a big problem (Liao and Ji, 2009). In this situation, eliciting knowledge from the domain experts is the feasible option.

When developing a BN for a specific domain, the knowledge from the domain experts is incorporated and used for constructing a model that is sufficient enough to represent the problem features (Flores et al., 2011). The knowledge is used to build the BN structure including the nodes, their states and the arcs between nodes. This is followed by parameterisation and then evaluation (Marcot et al., 2001; Cain, 2001).

2.4.1.2 Eliciting the CPT for a BN

CPT elicitation is commonly known as the most time consuming process in the BN elicitation (Druzdel and Van Der Gaag, 2000). This is due to the size of CPT that could be very large depends on the number of parents and states. When eliciting

CPTs using data, a large amount of historical data is needed, while using expert knowledge to elicit the CPTs, large amounts of time is required.

In this research project, eliciting the CPTs using historical data is not an option due to the lacking of valid and accurate historical data (described in Subsection). Therefore, eliciting CPTs using experts' knowlegde is more suitable in this research. When the number of CPTs to be populated is small, direct assigment method is mostly used (Liu et al., 2015). In this method, experts could easily estimated the probability value. However, when the number of CPTs are large, asking experts for directly providing the answer could put too much pressure on the experts (Das, 2004). Fuzzy logic and weighted sum algorithm are two approaches that commonly used to populate a large number of CPTs. Liu et al. (2015) employs fuzzy logic to construct less heuristics rules elicited from the experts and infer more CPTs. In the weighted sum algorithm proposed by Das (2004), experts estimates less CPs. The number of questions that need to answer by expert could be reduced and thus the elicitation effort. In this thesis, the approach proposed by Das (2004) to populate the CPTs. There are three distinct phase provided in this approach to populate the CPTs. The first two phases are used to generate the questions and the third phase is used in populating the CPTs.

1. Eliciting compatible parental configuration (CPC)
   In this phase, the experts are asked to specify the set of states in a parent node that are compatible to the state in other parent nodes. Consider the network in Figure 2.10, over the child node E, experts are asked to specify the parental combination of parent node PM, PT, and ME. For example, the experts start from the parent PM and interpret the compatible parental configuration as follows:

$$\begin{aligned} &\{Comp(PM = s)\} \equiv \{Comp(PT = s)\} \equiv \{Comp(ME = s)\} \\ &\equiv \{PM = s, PT = s, ME = s\}, \text{for} s = vl, l, a, h, vh \end{aligned} \tag{2.8}$$

   This configuration means there is equivalent relations of parent node PM in a specific state with parent node PT and parent node ME in the same specific state, for all the available states.

2. Eliciting conditional probabilities for each CPC
   Based on the compatible parental configuration provided above, the probability

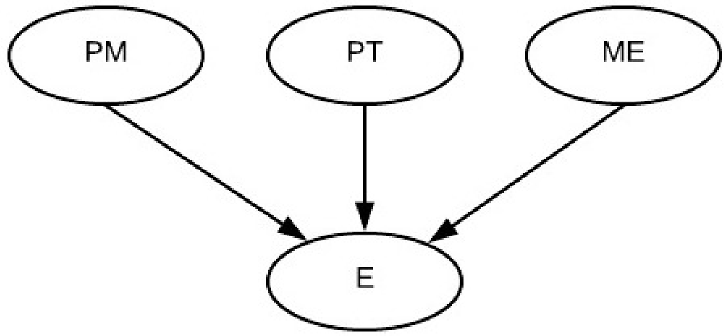


Figure 2.10: Example of a network with multiple node

distribution over the child node E will have:

$$p(E = e|\{Comp(PM = s)\}) = p(E = e|\{Comp(PT = s)\}) \\ = p(E = e|\{Comp(ME = s)\}), fore, s = vl, l, a, h, vh \tag{2.9}$$

The equation means that if the experts provide five probability distributions over the node E corresponding to the parental configurations $\{Comp(PM = s)\}$, $s = vl, l, a, h, vh$, then all the distributions for compatible parental combination are obtained.

3. Eliciting the relative weights
   A relative weight value (between zero and one) for each parent node, denoting the degree of influence a parent has on a child node (Baker and Mendes, 2010). The sum of relative weights for all parents of a child node should equal to one. If the relative weight of a parent node equal to zero or closer to zero, it means that parent node has no or less influence to the child node. If the relative weight of a parent node is equal to one or closer to one, it indicates that a parent node is the determinant of the condition probabilities in the child node.

2.4.2 Reasoning BNs with Number

BNs can be used to produce reasoning in relation to a specific domain. The observation of values in some nodes can be conditioned with new information. The process of conditioning is also known as probability propagation, inference or belief updating (Pearl, 2000). The idea is to provide the network with new evidence and analyse current beliefs in order to predict a result or identify a cause. There are four types of qualitative reasoning that are supported by BNs (Korb and Nicholson, 2011): 1) diagnostic reasoning; 2) predictive reasoning; 3) intercausal reasoning; and 4) mixed reasoning.

The quantitative reasoning using the actual number also can be done for BNs. The prior belief for value in each node or known as prior probability distribution (P) can be computed before any evidence is inputted to the network. In order to understand the process of using exact number in BNs reasoning, let use the example of lung cancer showed in Figure 2.11. The prior probability distribution of *Smoker* node is set to 0.3 (P(S) = 0.3). If there no evidence is inputted, the conditional probability (P(X|Y)) and prosterior probability distribution (Bel(X)) are originally specified in table in Figure 2.11. When evidence is inputted in any of the nodes, the (Bel(X)) value is updated, as shown in the first set of Table 2.2. If the smoking rate in the population increases to 50%, then the prior probability distribution of *Smoker* node is set to 0.5 (P(S) = 0.5). The second set in Table 2.2 shows the updated belief for non-evidence and if there is evidence inputted into the nodes. From these two set of updated beliefs it can be seen that when the evidence is about the patient being smoker (S=T), the prior probability distribution (P(S)) becomes irrelevant, because both networks give the same number.

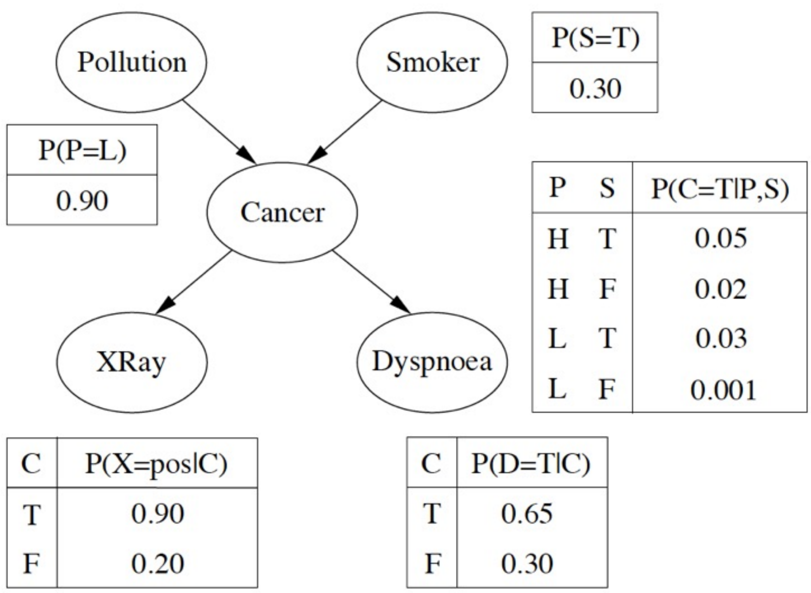


Figure 2.11: Example of BN for lung cancer problem (Korb and Nicholson, 2011)

Table 2.2: Updated beliefs given new information with smoking rate P(S)=0.3 and P(S)=0.5. Source: Korb and Nicholson (2011)

Node	**No Evidence**	**Reasoning Case**			
		Diagnostic	**Predictive**	**Intercausal**	**Combine**
P(S)=0.3		D=T	S=T	C=T & S=T	D=T & S=T
Bel(P=high)	0.100	0.102	0.100	0.156	0.102
Bel(S=T)	0.300	0.307	1	1	1
Bel(C=T)	0.011	0.025	0.032	1	0.067
Bel(X=pos)	0.208	0.217	0.222	0.900	0.247
Bel(D=T)	0.304	1	0.311	0.650	1
P(S)=0.5					
Bel(P=high)	0.100	0.102	0.100	0.156	0.102
Bel(S=T)	0.500	0.508	1	1	1
Bel(C=T)	0.174	0.037	0.032	1	0.067
Bel(X=pos)	0.212	0.226	0.311	0.900	0.247
Bel(D=T)	0.306	1	0.1	0.222	1

2.5 Combining Multiple Responses

2.5.1 Weighted Mean

The most commonly used method of combining a set of answers is to calculate a single summary value based on all the values in the dataset (Meyer and Booker, 2001). Mean or the average of the values in the dataset is one of the popular way to calculate the single value. However, by only calculating mean and using that number could possess a serious implication. For example, if the number of experts is small and one expert provide an answer that is far away from the value provided by the rest of the experts, then this extreme value greatly influence the mean value.

To overcome the influence of extreme value, some analysts prefer to use a weighted mean. Each expert answer is given its individual weight and the mean is calculated (see Equation 2.10). Meyer and Booker (2001) mentioned that the advantages of this method is the analyst could control the value or experts that provide more influence to the single value.

$$weightedmean = \sum_{i=1}^{n} x_i w_i / \sum_{i=1}^{n} w_i \tag{2.10}$$

However, determining the weight is not an easy process. A variety of methods can be used to determine weights, one of them is Saaty pairwise comparison or known as Analytic Hierarchy Process (AHP) (Saaty, 1990). The advantages of using AHP is the whole number of comparisons can be reduced via a hierarchy structure and the consistency of responses verified via a consistency ratio. In this research, experts with different background and knowledge are involved and is necessary to ensure the weight of each experts is well-assigned. Therefore in this research, AHP is the method that chosen to calculate the weight.

2.5.2 Analytic Hierarchy Process (AHP)

AHP is a multiple criteria decision-making approach and was introduced by Saaty (1990) to guide decision makers rank information based on pairwise comparison of two criterias or variables. In order to do pairwise comparison, the information that is needed includes:

- Which criteria or variables are more important compared to others
- The ratio of how much one variable is preferred over the other. The ratio can be represented by number between 1 to 9.

Generally, AHP follows three major steps (Saaty (2008)):

1. Establish the hierarchy of the structure
   The first step in the AHP is to model the problem as a hierarchy. There are three major levels contained in the hierarchy. The first or top level represents the overall goal of using this AHP method. The intermediate level represents the criteria that contribute to the goal and the bottom level are the objects. In this research project, the overall goal is to obtain the weight of each expert. While the bottom level are the experts which have given the weight.

2. Elicit the pairwise comparison judgments
   In the pairwise comparison, the node at each level will be compared, two by two, with respect to the nodes above them. The nodes in the criteria will be compared in pair with respect to the goal node, while the nodes in the alternatives level will be compared with respect to each node in the criteria. The scale that is used for the comparison judgement followed the fundamental scale proposed by Saaty (1990) (See Table 2.3). Once the weight is assigned, the next step is to transfer the weights to a pairwise comparison matrix and calculate priority vectors. The vector of priorities is the principle eigenvector on the matrix.

3. Establish the composite or global priorities of the lowest level with respect to the goal
   Once the priorities of the Criteria with respect to Goal, the priorities of

Alternatives with respect to the Criteria, the priority vector of Alternatives with respect to Goal can be calculated.

Table 2.3: Fundamental scale for the pairwise comparison

Intensity of importance scale	Definition	Explanation
1	Equal importance	Two elements or criteria contribute equally to the goal
3	Moderate importance of one over another	An element or criteria slightly favour over another based on the experience and judgment
5	Essential or strong importance	An element or criteria strongly favour over another based on the experience and judgment
7	Very strong importance	An element or criteria is strongly favoured and its dominance demonstrated in practice
9	Extreme importance	An element or criteria is extremely importance or preferred over another
2, 4, 6, 8	Intermediate values between the two adjacent judgements	When compromise is needed

2.6 Chapter Summary

This literature review introduced the peatland fire problems in Indonesia. It covered the possible contributing factors and the impact of fires to humans and the environment. The review then looked at how the fire management system in Indonesia deals with these fire problems to address RQ. 1 in Section 1.4. This included what can be learned from the fire management system that is currently implemented in Indonesia. This review found that the most of the current approaches implemented in Indonesia have not considered the unique characteristics of peatland fires such as human involvement and the characteristics of fuel. The review then expanded its scope to include the forest fire prediction tools that implemented in other regions. However, due to the different characteristics of fire, implementing the fire prediction tools from other regions is not suitable for Indonesia.

Learning the behaviour and characteristics of peatland fires is the first step to enhancing the peatland fire management system in Indonesia. The review looked at knowledge elicitation as part of the process to learn about the behaviour of peatland fires. The knowledge elicitation process uses focus groups and surveys, which was also discussed in this chapter as the method to gather the domain knowledge from an expert. This expert knowledge is presented in a graphical model, thus a brief introduction on BNs as graphical modelling tools was introduced in this review.

Chapter 3

Research Methodology

Introduction

This research project is framed using the Design Science Research (DSR) methodology outlined in Peffers et al. (2007). The chapter begins by briefly introducing both the DSR methodology and the reason for using the DSR methodology. The following section identifies the research problem, followed by the artifact requirements and the evaluation. The artifacts produced in this research are aimed at achieving the objective of this research. This chapter also explains the validation of the use of design science research in this research.

3.1 Design Science Research Methodology (DSRM)

Design science is a problem-solving paradigm. It is the attempt of researchers to develop an innovative artifact that can help solve problems (March and Smith, 1995). Design science research involves the process of creating artifacts to solve a problem, providing a clear contribution to the research, evaluating the design, and communicating the results to appropriate audiences (Von Alan et al., 2004). The artifacts can be in the form of constructs, models, methods, and instantiations. When there is existing artifact already available, design science can be used to improve the performance of the artifact to address the problems faced by human beings (Wieringa, 2009).

Peffers et al. (2007) proposed a design science research methodology (DSRM) as a framework that can be used in design science research. This methodology incorporates the principles, practices, and procedures required to support the objective of the research. DSRM requires an iterative procedure for the development of the artifacts for problem solving. The design of artifacts is only completed when they satisfy the requirement and constraints of the problem that need to be solved. However, there is always a possibility that the research will return to the development stage to improve the effectiveness of the artifacts (Von Alan et al., 2004; Peffers et al., 2007). This iterative procedure becomes one of the reasons for choosing DSRM as the research methodology in this research project. As shown in Figure 3.1, this research also involves the iterative process to confirm that the artifacts have fulfilled the requirements of the problem-solving. The iteration can be performed after the demonstration and/or the evaluation process. The outcome of the demonstration and the evaluation process is expected to show whether the artifacts are ready to be implemented or still need some further adjustment. If further adjustment is needed, it is always possible to return to the design step to refine the causal model or even refine the requirements for the final artifacts.

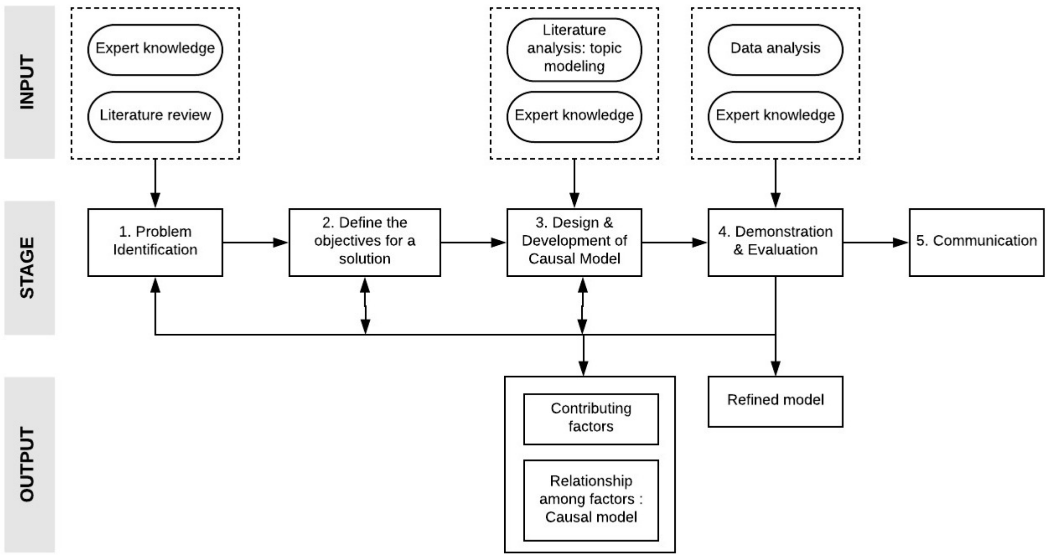


Figure 3.1: DSRM Process Model

Different sources of information and knowledge such as literature and expert knowledge are incorporated in almost every stage of the process. In the first stage of the DSRM process, a comprehensive literature review and interviewing of experts is conducted to gather information associated with the problem of peatland fires in Indonesia. In this stage, officers and experts were interviewed. The officers are from the local fire authorities in Central Kalimantan. The fire experts are involved in fire projects in Indonesia, including people from the Kalimantan Forest and Climate Partnership (KFCP). The purpose of this interview stage is to gather further information on the peatland fire problem based on the real life experience of the experts. Information from the literature and knowledge from the human experts is also incorporated into the third stage of the DSRM in order to design and develop the causal model. A collection of documents in the form of journal articles, conference papers, reports, and news articles is gathered and analysed. A literature analysis of the topic modeling method was conducted to identify the causal variables of the model. The fire experts involve in the knowledge elicitation process are chosen based on the criteria of education background, expertise, and practical knowledge in dealing with peatland fires. One way to identify these experts is based on the works found in the literature. For example, one expert was chosen because he published many papers explaining the causal relationship of human involvement in peatland fires. Other experts were chosen based on their local knowledge. As people who lived in fire incident areas, they are expected to be able to share their local wisdom and experience. Some experts were recruited based on recommendations from the selected experts. The same group of experts are involved again in the fourth stages of the DSRM to evaluate the causal model. A summary of the definition and scope of each source involved in the DSRM is presented in the table 3.1.

The present research project has been granted an ethical approval from the Monash University Human Research Ethics Committee (MUHREC) dated 31 October 2017. The letter of approval is presented in the Appendix A .

3.2 Problem Identification and Motivation

The DRSM process starts with the identification of the research problem. At this stage, the problem's purpose and the importance of its solution are presented (Peffers et al., 2007). The identification of the problem can be concluded through information gathered from the literature. Experts' knowledge can be used to determine the relevance of solving the problem (Offermann et al., 2009). In this research (see Figure 3.1), the problem related to peatland fires in Central Kalimantan is identified through literature review and conversation with fire experts. The following subsections discuss the problem with the current Indonesian fire management system and what is needed to overcome the problem. The scope of this research and goals are also identified in this section.

3.2.1 Problems with Management of Indonesian Fire

Peatland fires in Indonesia, especially in Central Kalimantan as explained in the literature review Section 2.1, have caused a significant economic, social, health and environmental cost not only for Indonesia but also the neighbouring countries (Page and Hooijer, 2016). The Indonesian government has drawn up a range of policies and regulations to minimise these fire problems. However, the threat of future fires seems to be continuing (Herawati and Santoso, 2011; Saharjo, 2016). The regulations seem

Table 3.1: Definition of literature and experts

DSRM Stages	**Input**	**Definition**	**Scope**
Stage 1	Literature Review	A comprehensive review on the literature related to forest and peatland fires in Indonesia	To identify the problem of peatland fires in Indonesia.
	Expert knowledge	Officers from the fire authorities in Central Kalimantan and Researchers from KFCP.	To gather further information and understanding of the peatland fire problems based on real-life experiences.
Stage 3	Literature Analysis: Topic Modeling	A literature analysis using a topic modeling method on the literature related to peatland fire occurrence in Indonesia	To identify factors contributing to peatland fires and use it as the causal variables of the model.
	Expert knowledge	Fire experts were chosen based on few criteria: (1) expertise; (2) experience with peatland fire problems; (3) local knowledge.	Through a focus group discussion and survey, expert knowledge was elicited to create the peatland fire causal model.
Stage 4	Data Analysis	A literature analysis using a topic modeling method on the literature related to peatland fire occurrence in Indonesia	To quantify the expert knowledge and present it in a quantitative manner.
	Expert knowledge	The same group of experts that were involved in Stage 3.	To evaluate the structure and result of the causal model.

to prioritise the suppression and the emergency response when fires occur rather than the prevention of fire outbreaks (Adinugroho, 2005). The prevention of fire outbreak is the early activity of fire management system that could reduce and minimise the damage and loss arising from fires (Martínez et al., 2009).

Preventing the occurrence of peatland fire in Indonesia is not an easy task, due to the complex and unique characteristics of the fires. Since peatland fire in Indonesia is mainly human-made (Harrison et al., 2009), the human to non-human interaction should be taken into account in the development of fire prevention systems. However, the current implemented fire prediction systems that are used by the Indonesian fire authorities do not consider this human involvement. For example, the fire danger rating used by the Bureau of Meteorology of Indonesia considers climate condition and vegetation-fuel as the only contributing factors in the current system (De Groot et al., 2007). The lack of consideration in addressing the underlying cause of peatland fires in Indonesia has caused ineffectiveness in the implementation of the current system used to tackle this fire problem (Herawati and Santoso, 2011).

A common method of detecting the occurrence of peatland fires in Indonesia is using hotspot information detected by satellites. These detected hotspots are used as an indicator of the occurrence of fire. However, not all detected hotspots are fires. Sometimes, it can be the sunglint from the water or the heat anomalies from coals underneath (Giglio et al., 2008; Allison et al., 2016). Thus, a ground investigation is always conducted to verify the information about the detected hotspots (Vayda, 2010; Saharjo, 2016). This method has been proven ineffective and inefficient because the cost is expensive and time-consuming, especially if the hotspots occur in inaccessible areas (Saharjo, 2016).

Having a better understanding of the characteristics of peatland fires could provide reliable information to find a solution to prevent the occurrence of peatland fires. This reliable information can support the decision maker when tackling fire problems (Applegate et al., 2002; Dennis et al., 2005). However, for Indonesian peatland fires, there is insufficient valid historical data that can be used to learn the characteristics of these fires.

3.2.2 Scope and Definition of Peatland Fire Escalation

The previous section briefly explains the problem in tackling the peatland fire problem in Indonesia. Peatland fires have a unique characteristic. Once a fire has ignited in a peatland area there are two possible outcomes: it can escalate to a large fire; or be extinguished because there is not enough fuel available. When the fire escalates, it can spread along and can burn the underground peat (Usup et al., 2004). Once the fires burn the peat underneath the forest, the emissions and pollutants resulting from this fire can affect human health. Therefore, it is important to prevent the fires from escalating to the surrounding area.

In this research, hotspot information as the starting point of peatland fires is used. (Giglio et al., 2008) mentioned that not all detected hotspots will translate into full-scale forest/peatland fires. However, there is a strong possibility that any hotspot turns or spreads into a fire. Therefore, Saharjo (2016) suggested identifying detected hotspots that can potentially turn into peatland wildfires. Through reliable information on hotspots that could escalate into peatland fires and their locations, the preventive actions can be focused on that area and taken immediately.

During the discussion conducted on 14 February 2017, experts suggested the definition of fire escalation as the probability of a hotspot escalates into peatland wildfire large fires on the following day. The Indonesian National Institute of Aeronautics and Space (LAPAN [1]) identifies a forest fire by the occurrence of multiple spot fires on the following day within a certain radius of the original spot fire (Roswintiarti et al., 2016). From these two definitions, this research defines fire escalation as the spread of spot fires detected in the next day beyond the initial spot fire(s) within the area of 2 x 2 km^2.

3.2.3 Problems in Gaining a Comprehensive Understanding of the Characteristics of Peatland Fire Occurrence

Understanding the relationship inherent in human to non-human interactions is an early step to dealing with the complex problems of peatland fire in Central Kalimantan, Indonesia. However, the problems that arise in obtaining reliable information about the condition of this type of fire are insufficient historical ground truth data and the consistency of human expert knowledge. Since the information on peatland fires is mostly available from narrative sources such as literature and human experts, this research project aims to explore these available resources to learn the characteristic of peatland fires. Therefore, a workflow is being designed that incorporates the information from literature and knowledge from human experts to learn the causal relationship among contributing factors in peatland fire occurrences.

The research questions that were formulated to address the problem above elaborated into three sub-research questions as explained in Chapter 1, Section 1.4.

[1]Lembaga Penerbangan dan Antariksa National

3.3 Defining the Objectives of a Solution

Once a problem is defined, it is important to determine how the problem should be solved. The objective of the solution can be quantitative and/or qualitative. An example of quantitative objective is measurement of the artifacts' performance. An example of qualitative objective is description of how an artifact supports solving the identical problem (Peffers et al., 2007).

The key problems as identified in Section 3.2 is the lack of sufficient historical data to learn the characteristics of peatland fires. The lack of data in Central Kalimantan has become a hurdle in learning a fire's characteristics using a data-driven approach. This complicates the process of predicting the escalation of hotspots to peatland fires. Therefore, other available resources are explored with an aim of learning the characteristics of these fires. Since the information on peatland fires is mostly available in the form of narratives and qualitative data, gathering information from these resources is a solution that can be implemented to solve the problem. The objective of this solution corresponds to the development of a workflow for incorporating the information from literature and knowledge from experts, to produce the causal model for predicting the escalation of hotspots into surface peatland fires.

There are three requirements artifacts have to meet in order to satisfy the problem specification:

a. Requirement 1 - Capability to integrate different forms of knowledge. In this research, knowledge was identified in the forms of quantitative and qualitative data. Literature and observation data provides quantitative data; the experts' knowledge provides qualitative data. The proposed workflow should be able to incorporate qualitative experts' knowledge and present the result in a quantitative manner.

b. Requirement 2 – Capability to accommodate the variability of the expert's experience and the language barrier. Having expertise from different experiences and knowledge about the forest fire could create a knowledge gap in the technical jargon that is used in the knowledge elicitation process. Experts can have different interpretations or cannot fully articulate the question. Another problem with having experts with different backgrounds is the difficulty of keeping the direction of the elicitation process focused to answer the problem.

c. Requirement 3 – Capability of the causal model in capturing the behaviour of peatland fires in Central Kalimantan. Peatland fires in Indonesia, especially in Central Kalimantan are mostly anthropogenic (i.e human-made) (Page et al., 2002). To capture the human contribution to the escalation of peatland fires, the causal model needs to involve people's viewpoints and not only the fire's viewpoint.

3.4 Design and Development of the Causal Model

Design science is a problem-solving paradigm (Von Alan et al., 2004) that produces an artifact to solve the problem (March and Smith, 1995). In principle, design science research tries to construct an artifact for a specific purpose and evaluate the performance of the artifact. The artifact should be able to address the unsolved problem. It can be a novel and innovative artifact or even the enhancement of an existing artifact (Von Alan et al., 2004). The design artifacts are classified into four types (March and Smith, 1995; Von Alan et al., 2004), (1) construct or concept, the conceptual vocabulary or symbol that is used to define the problem in the domain; (2) model, the proposition or statement expressing the relationship among constructs; (3) method, a set of steps to perform a specific task; and (4) instantiation, the implementation of a system. The process to create the artifact is part of the design and development process (Peffers et al., 2007). This activity includes defining the functionality of the artifacts, the architecture, and the construction process.

3.4.1 Deliverable Outcomes

There are two deliverable outcomes in this research project. The first outcome is a development framework for eliciting knowledge from literature and experts to construct a causal model. The second outcpme is a causal model to predict the escalation of peat fires to forest fires in Central Kalimantan.

a. Workflow to produce a causal model for fire escalation in peatland

   The first deliverable outcome produced by this research project is a repeatable workflow of incorporating the expert's knowledge to build a causal model for fire escalation in peatland. This workflow contains a set of stages and steps required to extract the information from literature and elicit knowledge from fire experts. This workflow can be used to quantify the causal relationship amongst factors that contribute to the escalation of peatland fires.

b. Causal model of peatland fire escalation to wildfire in Central Kalimantan, Indonesia.

   The second outcome is a causal model that represents the contributing factors and their relationship to peatland fire escalation. The model is developed using the proposed framework. Through the causal model, reliable information about surface fire can be acquired and used to deliver a preventive decision to minimise fire occurrence.

3.4.2 Workflow of Causal Model Development

The development of the causal model follows four stages (Druzdel and Van Der Gaag, 2000; Korb and Nicholson, 2011): i) identification of the causal variables; ii) identification of the relationships between variables and development of the causal model structure; iii) parameter estimation; and iv) model evaluation. A knowledge-based approach is applied throughout the development of the causal model. This approach incorporates the process of extracting information from literature and eliciting knowledge from human experts. The literature analysis is used to elicit the initial variables. The result from the literature analysis is used to develop the structure of the causal model. A focus group discussion is conducted to refine and evaluate the structure of the causal model. A workflow is presented for the knowledge-based approach in developing the causal model. This framework has four stages (see Figure 3.2): i) initial variable elicitation; ii) structure development; iii) parameter estimation; and iv) model evaluation.

3.4.2.1 Stage 1: Initial Variables Development

The first stage in the development of the causal model is defining the initial variables to be used in the model. The variables can be derived from the historical data, an existing conceptual model or even human judgement (Marcot et al., 2006; Pollino and Henderson, 2010; Korb and Nicholson, 2011). However, in some problem areas, no adequate data or models are available to elicit the variables. Meyer and Booker (2001) suggested that when the data from experiments or observations are limited, the acquisition of human knowledge can supplement the information. Therefore, this research project proposed an alternative way to identify the variables in the causal model. A literature analysis was carried out using text mining analysis to identify the initial variables. A hidden pattern in the literature was observed to gather topics and terms related to factors contributing to the peatland fire escalations.

A topic modeling technique using LDA algorithms is implemented to extract topics and terms from a a set of documents. These topics and terms will be interpreted as the initial variables in the causal model. The first step is collecting the documents. The documents can be in the form of journal articles, conference papers, reports, or news articles. After a set of documents is determined, the next step is to pre-process the document into a form the fits the topic modeling requirement. The standard procedure is to tokenize the documents and create bags-of-words.

Once the topics and terms are extracted using the LDA algorithm, these terms then interpret and identify the causal variables for the model. To make sure that the topics and terms resulting from this topic modeling approach can capture the characteristics of peatland fires in Indonesia, fire experts are asked to provide a list of contributing factors of peatland fire occurrence. The list of terms resulting from topic modeling are compared with the experts' opinions. Then the similarity of these lists are measured.

3.4.2.2 Stage 2: Structure Development

The second stage is structure development using the knowledge from experts. The expert knowledge elicitation process is conducted through a focus group discussion. The discussion involves experts with different backgrounds and expertise as described in Section 3.1.

This second stage starts with the refinement of the causal variables extracted from the previous stage. Questions that can be used as the guideline in this refinement process are: *Are the nodes in the structure the right ones? Are they named usefully?*. Once the experts agree on the causal variables, the discussion continues with the determination of the relationship within each variable. The structure of the causal model follows the graphical structure of Bayesian Networks (BNs) as described in Section 2.4.1. In the structure nodes represent the causal variables; direct arcs represent the qualitative relationship between the nodes.

To identify the relationship of each variable, each expert was asked whether the presence of one variable influenced the other variables. Since multiple experts were involved in the knowledge elicitation process, the answers should be combined to create a single answer. There are a few techniques to collate answers from multiple experts, such as majority voting and the expectation–maximization (EM) algorithm (O'Hagan et al., 2006). The majority voting technique combines the experts' answers and creates a single structure. This new structure is then presented to the experts, and another round of discussion is conducted to refine the structure. Questions such as *are the direction of the arcs right?* are used in this refinement process. The discussion also covers how to solve the variables with a looping problem, restructure the causal model based on compactness and node ordering rules, and generate the state for each variable. After the structure of the model is developed, the next step is to define the states of each node.

3.4.2.3 Stage 3: Parameter Estimation

The process in Stage 3 aims to quantify the relationship of each variable that was defined in the previous stage. As the graphical structure of BNs is followed, the relationship of each node is described in a conditional probability table (CPT). The entries in a CPT can be 'parameterised' using a range and combination of methods, including directly observed data, probabilistic or empirical equations, results from model simulations, or elicitation from expert knowledge (Pollino and Henderson, 2010). In this research, expert knowledge is mainly used to parameterise the CPT. The expert elicitation process in this stage is conducted through an online questionnaire. A set of questions is generated to gather the probability of each combination of states. Since the CPT is used to describe the probability within each state in relation to the states in other related variables, the size of the CPT can be very large. This leads to a large number of questions that should be asked to elicit the probability. To reduce the number of questions to be asked and thus the elicitation effort, the relative weight and compatible probability method proposed by Das (2004) is used to populate the CPT.

The online questionnaire was set up using a cloud-based online survey, SurveyMonkey. The invitation to participate in the survey was sent to the experts' email. The experts were given a specific time period of two weeks to finish the questionnaire. They did it in their own time to allow them to reflect and revise the answers. Once all experts completed the questions, the next step was collating the answers into a single probability answer. Each expert's answers are recognised based the expertise of the expert itself. Thus, a Analytical Hierarchy Process (AHP) (Saaty, 2008) was used to weight the experts' answers. In this research, the aggregation of probability distributions used the linear opinion pool approach (Clemen and Winkler, 1999):

$$p(\theta) = \sum_{i=1}^{n} w_i p_i(\theta)$$

where n is the number of experts, $p_i(\theta)$ represent expert i's probability distribution for unknown θ, $p(\theta)$ represents the combined probability distribution and the weights w_i are non-negative and sum to one.

3.4.2.4 Stage 4: Model Evaluation

The last stage of the workflow aims to evaluate the causal model to ensure that the characteristics of peatland fires can be presented. This evaluation process also checks whether the causal model can be used to predict the escalation of hotspots into surface peatland fires.

A comparative evaluation is conducted to observe how well the causal model predicts the escalation of hotspots to peatland fires. The result from the causal model is compared with the ground truth data and the result from the generated rules of the Indonesian National Institute of Aeronautics and Space (LAPAN [2]). The ground truth data is a set of hotspot data that has been verified as fire escalation based on the ground investigation conducted by the KFCP Fire Management Team (FMTeam).

Ground Truth Data

The dataset of fire escalation was obtained from the KFCP Fire Management Team (FMTeam) investigation. During the fire season in 2012, the FMTeam conducted a field investigation on the fire incidences on peatland areas. The purpose of the investigation is to gather further information on how and why fires occur on peatland. The investigation was directed by data collected from the hotspot satellites. The FMTeam rely on the daily hotspot data provided by MODIS satellites (Terra and Aqua). Due to the limitation of logistic (budgets, access, and time available), not all hotspots were investigated (Graham et al., 2014). Investigated hotspots were selected according to the following criteria:

- When several hotspots were clustered at single location, only a single point within a cluster was investigated.
- When hotspot occurred within the managed areas of the local communities, especially cleared land, only a proportion of the hotspots were investigated.
- Hotspots occurring in deep peat were investigated.
- Hotspots occurring on planting sites and community assets were investigated immediately
- Hotspots occurring on mineral soil being cleared with deliberate burning were not investigated

The data collected from the investigation are fire location, start/end date, land tenure, weather conditions, the cause of the fire, the motivation for the fire, the total area burnt, fire intensity, and fire damage.

LAPAN's Rule

LAPAN is one of the agencies in Indonesia that are responsible for providing remote sensing data such as active fire data or known hotspots. As previously describe in

[2] Lembaga Penerbangan dan Antariksa National

Section 2.2, a web-based map has been released by LAPAN. This web-based map contains information about the locations of hotspot occurrence and the levels of confidence. Based on the information from this web-based map, the fire authorities could be warned of the possibilities of forest fire incidents. In addition to the web-based map, LAPAN also created a guideline on how to read and analyse the hotspot data to detect the forest fire occurrence (Roswintiarti et al., 2016).

For consistency, LAPAN's guideline is adopted to detect the fire occurrence:

1. If a cluster of hotspots is found in the same day, it can be assumed that fire is happening in that area. The classification proposed by Suwarsono and Vetrita (2014) is used to determine the fire risk level based on the density of hotpots in a certain area.

    - Low risk: if less than two hotpots are found per km^2
    - Moderate risk: if between two and five hotspots are found per km^2
    - High risk: if between six and ten hotspots are found per km^2
    - Extreme risk: if more than ten hotspots are found per km^2

    This research chose a minimum threshold of five hotspots found per km^2, as set in high risk, to set up the cluster.

2. If there is smoke detected around the location of hotspot, then there is high probability of fires. In their guideline, LAPAN mentioned processing images through calculation of Red Green Blue (RGB) pixels from satellite imagery to determine whether there is smoke in the location of the hotspot. However, there is no further explanation about how this image processing is conducted. In this research, a simple RGB image processing is conducted using the satellite imageries from MODIS satellite Terra/Aqua Combination of spectral channel 1 (0.65μM) representing red component, spectral channel 2 (0.86μM) representing green component and spectral channel 7 (2.13μM) as blue component is used to indicate the smoke location from forest fires.

    The combination spectral channel 1 - 2 - 7, representing the RGB false colour MODIS, was obtained from Corrected Reflectance (Bands 7-2-1) product of the MODIS/Aqua Surface Reflectance Daily L2G Global 1km and 500m available on https://worldview.earthdata.nasa.gov/. In order to determine whether a hotspot occurred around the smoke location, the result from this RGB image processing was overlaid with the hotspot data. For hotspot that located in the intersection with the smoke pixels is considered as fires

3. If a hotspot reoccurred in the subsequent days in certain radius. Due to the pixel problem, most likely it is difficult to find a hotspot that reoccurred at the same coordinate. Vetrita and Haryani (2011) found that if a new hotspot occurs within a 2 km radius in the three days after an earlier hotspot that hotspot can be assumed to be a reoccurring hotspot.

A sensitivity analysis is conducted to determine the result of the comparative evaluation. There two sets of test cases are generated: test cases for hotspot escalation into peatland fire and test cases for non-escalation of hotspots. These test cases are generated using the ground truth data from the KFCP database. There are eight possible outcomes (see Table 3.2). These outcomes can be categorised into optimistic results and pessimistic results. In an optimistic result, the result from the causal model aligns with the ground check data and LAPAN's rule, whether the hotspot will/will not escalate into peatland fires. The pessimistic result shows disagreement within the result from the causal model, ground check data, and LAPAN's rule.

This sensitivity analysis aims to identify nodes or variables that are most sensitive to changes (Laskey and Mahoney, 2000). The output of each case in hotspot escalation and non-escalation is reviewed to determine whether the causal models deliver reasonable results. The causal models can then be adjusted based on the result of these evaluations.

Table 3.2: Possible outcomes for the scenario-based analysis

Ground Truth Data	Escalation/Yes				Non Escalation/No			
Causal Model	Yes		No		Yes		No	
LAPAN	Yes	No	Yes	No	Yes	No	Yes	No

3.5 Demonstration and Evaluation

In design science research, evaluation is needed to observe and measure how well the artifact supports a solution to the model (Von Alan et al., 2004). The result of this evaluation activity could be used to show how the artifact is both applicable and useful (Sonnenberg and vom Brocke, 2011). During the evaluation process, it is always possible to iterate back to the previous process such as "design the artifact" or even back to "identify the problem".

In this research, the evaluation process occurs throughout the design process. Stage 1 evaluates the robustness of using the topic modeling approach in identifying the causal variables by comparing the result with the experts' opinion. In Stage 3, an elicitation review is conducted as the evaluation process. The elicitation review involves an overall review of node definitions, state definitions, and the relationships associated with each node. This process is conducted using the expert analysis to ensure the causal model has the right structure, is simple, is easy to understand, and captures the characteristics of peatland fires.

This evaluation activity is also part of Stage 4 in the workflow. It is conducted after the construction of the artifacts. Based on the nature of the artifacts, the evaluation takes into account the applicability of the causal model to real world problems, the robustness of the causal model, the ease of use, and the generality of the workflow used to develop the causal model. As mentioned in Subsection 3.4.2, the evaluation of the causal model, three types of evaluation are performed: case-based evaluation, comparative evaluation, and sensitivity analysis.

Comparative Evaluation

The comparative evaluation aims to determine the performance of the causal model in predicting the surface peatland fire escalation. A confusion matrix is generated to define the nature of causal model in predicting the fire escalation whether its more towards false positive or false negative result. This result also compared with the implementation of the LAPAN's rule to determine whether the causal model deliver better performance compared to existing guideline, which is LAPAN's rule.

Sensitivity Analysis

The sensitivity analysis is used to refine and simplify the causal model. It is also used to determine the parameters that need further research, to strengthen the knowledge base, and to test the robustness of the causal model as a problem-solving method.

1. Subjective sensitivity analysis.

   The purpose of this activity is to simplify the causal model. This activity involves the fire experts working with the model before and also an external expert that does not join the development of the model. A focus group discussion and also an email conversation are conducted to investigate the unimportant variables that can be discarded.

2. One at a time sensitivity measures

   This method is used to determine the most influential variables in the fire causal model. In this method, each variable is varied while holding the other variables fixed. The change in the model output will be quantified, at a time when one input variable changes and rest are constant. Using Moris' AOT method, the input variable will be grouped into three categories:

(a) an input variable with negligible effects

(b) an input variable with large linear effects without interactions

(c) an input variable having large non-linear and/or interaction effects

3.6 Research Methodology Validation

The principle in building an artifact in design science research is having knowledge and understanding of a design problem and its solution. Von Alan et al. (2004) proposed seven guidelines to conduct and evaluate good design science research. The research methodology in this chapter is evaluated against the guidelines and presented in the explanation below.

1. Guideline 1: Design as Artifacts

   The outcome of design science is an innovative and purposeful artifact (Von Alan et al., 2004). The aim of this research is to construct and evaluate the proposed artifacts. Due to the lack of historical data, the development of this causal model relies on the knowledge and wisdom from experts and local people. Therefore, in this thesis the first artifact are the workflow for incorporating the experts' knowledge to build a causal model for predicting the escalation of peatland fires. The second artifact is the causal model that comprehensively show the relationship of factors contributing to the escalation of hotspot into peatland fires.

2. Guideline 2: Problem Relevance

   Von Alan et al. (2004) defined a problem as the differences between the aim and goal of a system and the current status of that system. An effective artifact should be able to address the problem faced by the system. The workflow introduced in this research project should be able to analyse the information and knowledge gathered from the literature and experts. The analysis result can be used to address the problem of peatland fire escalation.

3. Guideline 3: Design Evaluation

   Evaluation of the artifact is an important component in design science research. There are several evaluation methods that are applicable for design science research suggested by Von Alan et al. (2004). Of the suggested evaluation methods, this research uses an observational, analytical ,and descriptive method to evaluate the artifacts.

4. Guideline 4: Research Contribution

   Effective design science research must provide a clear contribution in three different areas (Von Alan et al., 2004). It can be in the areas of the design artifact, knowledge construction and knowledge evaluation. The contribution should be determined based on novelty, generality, and significance. The contribution of this research is comprised of the artifacts themselves. This is the first time the interaction of human and non-human actions are modelled as contributing factors in peatland fire escalation. It also demonstrates that a pure knowledge driven approach can be used to create a model focused on solving a real world problem.

5. Guideline 5: Research Rigor

   In design science research, rigor addresses the way in which research is conducted including the construction and evaluation processes of the artifacts (Von Alan et al., 2004). In this research project, a combination of qualitative and quantitative methods is used to evaluate the artifacts. Expert opinion is used to review the result of data analysis on the causal model.

6. Guideline 6: Design as a Search Process

   Design science is an iterative process to discover an effective solution to a problem (Von Alan et al., 2004). As described in Section 3.1, in this research project, an iterative process is conducted to ensure that the proposed causal model can be used to gain better understanding of peatland fire escalation.

7. Guideline 7: Communication of Research

   Design science research must be presented in order to provide a benefit to practitioners and other audiences (Von Alan et al., 2004). The artifacts of this research project are presented in a peer-reviewed conference and journal article.

3.7 Chapter Summary

The purpose of this chapter was to develop and justify the methodology chosen for this research. Design Science Research is the paradigm that was chosen. It is a suitable approach to investigate problems in the domain of Information Technology. The research activities follow the Peffers's DSRM research model (Peffers et al., 2007) outlines to ensures a rigorous research process.

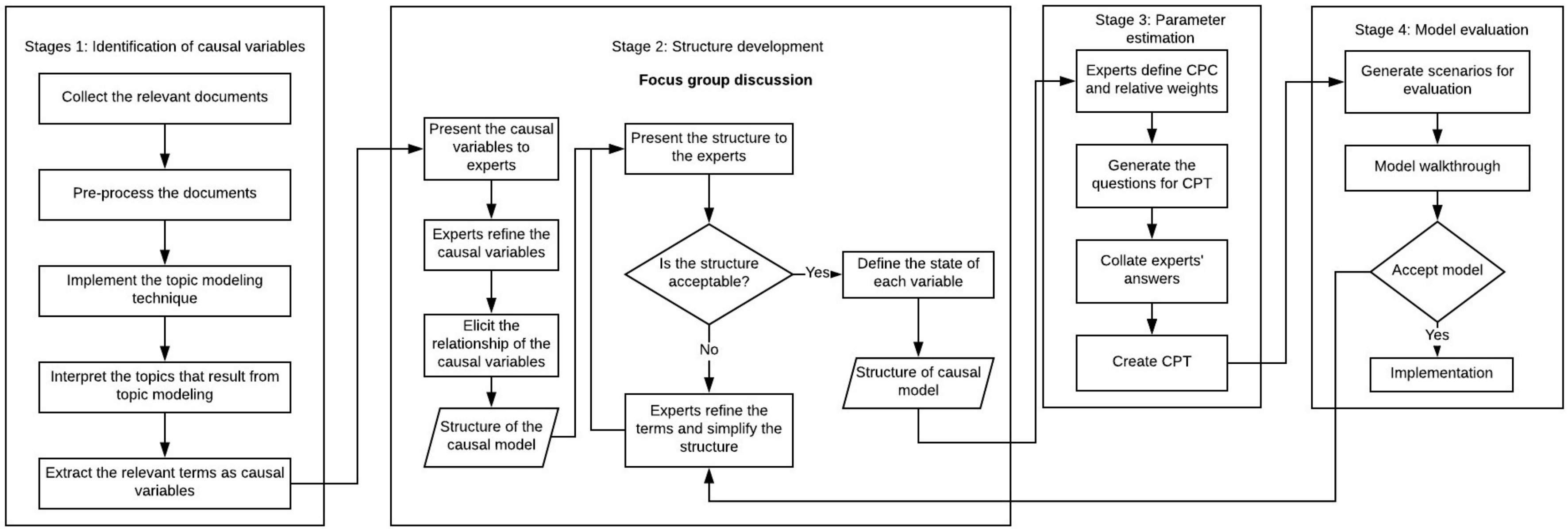


Figure 3.2: Workflow for incorporating the experts' knowledge to build a causal model for predicting escalation of hotspots to wildfires in peatland

Chapter 4

Automated Identification of Causal Variables

This chapter presents the findings from the process of extracting information in literature. The findings particularly address the second research question about 'how' and 'why' topic modeling techniques can be used to identify the causal variables influencing the hotspot escalation into peatland fires. This information extraction process is the first stage of the proposed workflow of causal model development (see Figure 3.2).

The structure of this chapter follows the steps of the first stage of the causal model development workflow (see Figure 4.1). It starts with collecting the relevant documents, described in Section 4.1. Then follows pre-processing the documents, described in Section 4.2 and implementing the LDA algorithm to extract the topics and terms, described in Section 4.3. In Section 4.4, the process of interpreting the terms into the causal variables is described. In Section 4.5, the terms extracted from the literature using the topic modeling method are compared with the experts' opinions. The chapter concludes with a summary of why the topic modeling technique can be used to identify the causal variables for hotspot escalation into peatland fire.

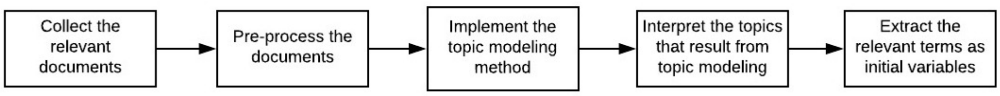


Figure 4.1: A generative process of discovering topics using the topic modeling technique

4.1 Collecting the Relevant Documents

The documents collected for the elicitation of knowledge are in the form of book chapters, journal articles, conference papers, and project reports. The journal/conference papers were obtained from the Scopus database and Google Scholar. The project reports were gathered from the Indonesian government and non-government organisations. For journal/conference papers, the Scopus database is the preferable resource. Most papers related to forest fires in Indonesia would be written by Indonesian researchers or involve Indonesian researchers as co-authors. The researchers in Indonesia are encouraged to publish their papers in journals or at conferences that are Scopus indexed. Therefore, many articles related to forest fires could be found in the Scopus database. The document searching is also expanded to Google Scholar database, in order to gather the papers published in journals or conferences without a Scopus index.

The search engines in the Scopus database and Google Scholar are optimised. Thus, complex search queries are allowed through the use of specific fields and Boolean operators "AND" and "OR". I used Boolean expressions containing "OR" connected expressions for the terms related to forest fire and location. Both were connected through "AND" expressions. For the terms related to forest fires, the query was not limited only to the term *forest fire* but also considered terms *peat fire* and *peatland fire*. This because the fires in Central Kalimantan mostly happened in peatland areas. The term *anthropogenic fire* was also included because the nature of forest fire in Indonesia is man-made. The term *fire behaviour* was also considered to expand the searching criteria. Terms related to the location are *Indonesia*, and *Central Kalimantan*. The query was run in the title, abstract, and keyword of the articles. The query was also limited to search documents that were written in English and published after 1995. The documents were book chapters, journal articles, or conference papers. The documents published in medicine, pharmacy, economy, business, and biochemistry domains were excluded. These domains were excluded because mostly discussed the impact of forest or peatland fires. While the aim of the document searching was to gather only the documents that focused on the behaviour of a forest/peatland fire, how it spreads, and what are the contributing factors. In total, a collection of 561 papers were obtained, which spanned 21 years from 1995 to 2016.

In addition to documents collected from the Scopus database, project reports from the government or non-government institutions were also included in the analysis. The institutions included the Ministry of Forestry, Center for International Forestry Research (CIFOR), Kalimantan Forests and Climate Partnership (KFCP) and World

Resources Institute (WRI). Ten project reports from different organisations and institutions were added to the collection.

In addition to documents collected from the Scopus database, project reports from the government or non-government institution were also included in the analysis. The institutions included the Ministry of Forestry, Center for International Forestry Research (CIFOR), Kalimantan Forests and Climate Partnership (KFCP) and World Resources Institute (WRI). Ten project reports from different organisations and institutions were added to the collection.

Research in topic modeling mostly analyses only the title or abstract of papers. In this thesis, the analysis was conducted on the whole content of the document, from the abstract to the conclusion. The purpose of analysing the entire content due to detailed explanations of the fire behaviour sometimes occurring in the introduction, discussion, and conclusion. The acknowledgement and reference parts were excluded because the content of this part sometimes gives irrelevant information. However, a small portion of the documents were found without full-text versions available. For these papers, only the abstract was included in the analysis.

4.2 Pre-processing the Documents for Terms Extraction

The quality of the analysis could be improved by taking a few steps to prepare the documents. The steps started with the extraction of terms from those filtered papers. Documents were split into individual words and sequences of words using whitespace characters. The punctuation, numbers, and non-word characters from the documents were removed. Terms also removed included stopwords, meaningless terms. Terms were stemmed as explained in the subsection below. The remaining words then were converted into lower-case characters to reduce the number of distinct word types caused by the lower and higher cases.

4.2.1 Removing stopwords

Stopwords usually refer to the most common words in a language and are non-significant to the content of a document or their information value is almost zero (Meyer et al., 2008). The stopwords were removed to reduce the noise in the textual data. Removing the stopwords can also be used to reduce the size of the term matrix without losing the significant relations inherent to the matrix (Wu et al., 2006). The pre-compiled stopwords list that exists in the literature (Rijsbergen,

1979) was used. In addition to that list, the most frequent words, single words, words that occur in fewer than four documents and some meaningless terms were added to the stopword list. The meaningless terms usually tell very little about the similarity of the documents. One type of meaningless term that frequently occurred in our document term matrix is the name of an author such as, *page*, *van*, *hooijer*, and *wooster*. These names frequently occurred because the collection of papers is particular for peatland fire in Central Kalimantan, Indonesia. Those researchers and their work have been cited many times in recent articles. Even after the reference list is excluded from our text analysis, these authors' names still can be found in the in-text citation.

4.2.2 Stemming

Stemming is used to reduce the different variants of word forms and transform them into common word roots (Xu and Croft, 1998). For example, *updating*, *updates* and *updated* have the same common root which is *update*. One of the processes in stemming is deleting the a word's suffix. The result from this process sometimes delivers incomplete words or ambiguous words. For example, the result of the stemming process transforms the word *example* and *examples* to an incomplete word form of *exampl*. *Update* and *updated* were stemmed into *updat*. These incomplete words might provide a confusion in the term interpretation process. To avoid the confusion and misinterpretation, the stem result should be restored in such a way to be close to the root word and the original. This stemming process will deliver words that look normal and complete, so it will be easy to understand. A dictionary which contains the words from the original documents was created. The words resulting from the stemming process are then compared with the words in dictionary.

4.3 Using Topic Modeling Techniques to Extract the Relevant Terms from the Literature

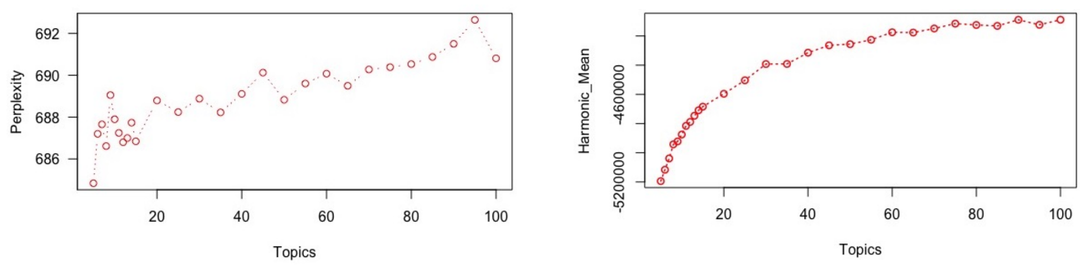
In this thesis, a topic modeling technique using the LDA algorithm was used to extract the topics and terms. The definition of topic modeling and the LDA algorithm can be found in Chapter 2, Subsection 2.3.1. In this Section, the LDA model is applied to a set of documents to elicit the terms that relate to the factors contributing to the escalation of a hotspot into peatland fire. Once all of the documents are converted into the correct format, the LDA algorithm is implemented to extract the topics containing the terms. Each topic is then interpreted based on the terms occurring in the topic. Relevant terms are chosen as causal variables.

Before extracting the terms and interpreting them as causal variables for peatland fire escalation, it is important to evaluate the performance of the model resulting from LDA algorithm. The evaluation process of the performance of the topic modeling involves a process to ensure the generalisation and interpretability capability of the model. In this thesis, to measure the generalisation of the topic models, the probability of unseen held-out documents of some training documents is estimated. In Section 4.3.1, a perplexity value was calculated to select a model with a better generalisation value. In Section 4.3.2, the topic coherence of the n-top words related to the topics was computed. The purpose of computing the topic coherence is to gather information about the ease with which the topics can be understood and interpreted.

4.3.1 Finding the Number of Topics

The LDA algorithm requires some basic input parameters, such as the number of topics K and the prior Dirichlet topic distribution. In this research, the number of topics K is not known yet. Therefore, evaluation calculating the perplexity and harmonic means was conducted in order to find the best number of topics.

The evaluation was conducted by estimating the probability of unseen held-out documents given some training documents. The inverse of log-likelihood (perplexity) (Heinrich, 2005) and harmonic mean (Griffiths and Steyvers, 2004) were used to measure the performance of a model to generalise the unseen data. In order to find a suitable number of topics, there is a perplexity calculation over the documents for a different number of topics T = 5 to T = 100 using 10-fold cross-validation on the training dataset. As suggested in Griffiths and Steyvers (2004), with $\beta = 0.1$ and $\alpha = 50/T$, 1000 Gibbs sampling iterations are used. The estimation of perplexity and harmonic mean is shown in Figure 4.2.



(a) Perplexity

(b) Harmonic Mean

Figure 4.2: Discovering the appropriate number of topics

Choosing the appropriate number of topics K for a set of documents is one of the keys to successfully applying the topic modeling. If the K is too low, it will generate topics that are overly broad, while the high K value can result in over-clustering the data. As shown in Figure 4.2(a), the lowest perplexity is found in the model with five topics. However, five topics is considered too few. A model with a small number of topics might contain a merge of topics that should remain separated. Other models were examined. Models with 8, 12, 13, and 15 topics also had quite low perplexity scores. One of these models could be considered to have an appropriate number of topics.

However, Figure 4.2(b) shows that the harmonic mean value increased as the number of topics was also increased. The higher harmonic mean value of a model means that the model should be accounted as the appropriate choice (Griffiths and Steyvers, 2004). From the measurement of harmonic mean it is difficult to decide the appropriate number of topics, since the harmonic mean value was increasing when the number of topic increased. When too many numbers of topics were used, this can be caused an overfitting. Therefore in this thesis, the number of topics that is resulted from perplexity measurement was considerable as the appropriate number of topics.

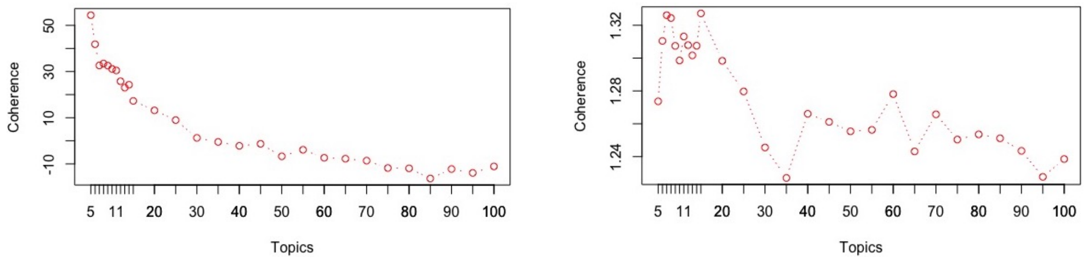
Having the appropriate number of topics does not guarantee that the topics will be easy to interpret. The results from perplexity and the harmonic mean do not reflect the semantic coherence of individual topics learned by a topic model. Therefore, the topic coherence for each model was measured to find out the best model to use. In the next subsection, the quality of the model is evaluated by comparing the topic coherence of single topics. The quality of the topics resulting from topic modeling can contain coherent topics that are easy to understand and interpret. But it also can have random and absurd topics that cannot be easily decomposed into individual topics.

4.3.2 Topic coherence

There is no guarantee that all of the topics generated by a topic model will be easy to understand and interpret. Topic coherence is used to measure the degree of semantic similarity between high scoring words in the topic (Stevens et al., 2012).

For all experiments, the coherence of each topic was computed based on the 10 words with the highest weight. The parameter was set to $\epsilon = 10^{-12}$, because both UCI and UMass perform better if the parameter ϵ is chosen to be small (Stevens et al., 2012). Before comparing the model, an aggregate measurement is conducted to evaluate the quality of the complete model instead of the individual topics. In this thesis, the average coherence of all topics is used to represent the summarization of the model's quality.

Figure 4.3 shows the average coherence scores for each model with different number of topics. Based on the UMass measurement, it is clear that learning more topics decreases the quality of the model. While the UCI score shows the different levels of topic quality when the number of topics is increased. Based on the UCI measurement, there are three models with the highest score, models with 7, 8, and 15 topics, while the UMass measurement shows that the quality of a model significantly decreased after the number of topics reached 20. Based on the manual evaluation of the word distribution in models with 7, 8, and 15 topics, the model with 15 topics provided a wide range of topic variety compared to having 7 and 8 topics. Thus, in the next experiments the model with 15 topics was used to gather the terms.



(a) UMass (b) UCI

Figure 4.3: Average topic coherence for each model

4.4 Terms interpretation of the causal variables

Based on the evaluation presented in Section 4.3, the model with 15 topics was chosen and used to generate topics and terms for the causal variables. The LDA model was run with 15 topics and obtained two types of posterior distribution. The first type is posterior topic distribution of each document; the second is posterior word distribution of each topic K.

4.4.1 Labeling the topics

Generating meaningful labels for topic word distribution can facilitate the process of interpretation of topics. The label can be manually generated using human interpretation or automatically generated using the existing method. The problem with manual labelling is that the labels generated are usually subjective and can easily be biased towards the user's personal opinion (Mei et al., 2007). Therefore in this thesis, an automated labeling method from Mei et al. (2007) was implemented to label the topics learned from the topic modeling method. The steps to generate the label can be found below:

1. Extract a set of candidate labels from a reference collection

   At first, a set of meaningful phrases from 2-word ngrams based on a statistical test was extracted. The significance of 2-grams was then tested using T-Test and only phrases with positive values were extracted.

2. Design a relevance scoring function

   The relevance scoring function will be used to rank the labels based on the semantic similarity with a topic model.

3. Rank the candidate label. Using the score, the candidate labels are ranked with respect to each topic model. Only top-ranked labels were chosen to label the corresponding topic.

The candidate labels for the 15 topics were ranked. A subset of the example topics is shown in Table 4.1. Each topic contains a list of words with the highest probabilities. It can be seen for some topics that it is difficult to describe the meaning of the topic merely from the top words. For example, Topic 7 and Topic 11 seem to be a similar topic, especially as the first two words with the highest probability are the same. However, based on the generated auto label, these two topics have different labels which lead to different meanings of topics in the interpretation process. From this experiment, it is shown that the automatically generated labels can capture the meaning of a topic to some extent, even though there are some confusing labels. The confusing label could be coming from some topics that are difficult to interpret.

Table 4.1: Sample of topics and the auto generated labels

	Topic 1	**Topic 3**	**Topic 7**	**Topic 11**	**Topic 15**
Auto Label	**climate change**	**el nino**	**land cover**	**land use**	**soil moisture**
theta	forest change tropic increase climate region global	rainfall nino season month data kalimantan drought	forest land plantation log cover kalimantan degraded	land forest cultivate village rubber communities local	soil site forest sample temperature flux depth

4.4.2 Discovering topics

Using the auto-generated label as guidance, the theme of each topic can be interpreted. For some topics, an auto-generated label could be adapted as a topic's label. However, for some other topics the auto label needs to be changed, to make it straightforward to the user. For example, an auto-generated label for Topic 3 is *el-nino*. After conducting a further investigation into the terms inside this topic, the label for this topic could be broadened to *weather condition*. This assumption was made based on the terms that occur in the topic such as *rainfall*, *drought*, *dry*, and *temperature* that most likely represent the weather conditions during the fire season.

The revised label for each topic and the top-20 terms inside the topics are presented in Table 4.2.

Table 4.2: The revised label and most 20-words for 15 topics.

Topic	**Auto label**	**Revised Label**	**Most 20-words**
1	**climate change**	**climate change**	*forest, change, tropic, increase, climate, region, global, asia, effect, deforestation, human, ecosystem, gap, dynamic, impact, regime, southeast, land, particular, agriculture*
2	**forest canopies**	**forest ecosystem**	*forest, species, tree, plot, burn, densities, studies, log, canopies, type, seed, disturb, significance, site, stem, unburn, composites, tropic, found, recoveries*
3	**el niño**	**weather condition**	*rainfall, niño, season, month, data, kalimantan, drought, period, dry, region, event, mean, enso, active, indonesia, fire, anomalies, day, borneo, temperature*
4	**carbon emission**	**carbon emission**	*model, estimates, emission, data, carbon, value, biomass, variable, burn, use, table, calculates, total, global, measurement, mean, includes, factor, spatial, product*
5	**secondary forest**	**forest condition**	*forest, manage, species, soil, ecology, indices, nature, protected, degraded, agriculture, conserves, water, communities, biodiversity, habitat, product, secondaries, ecosystem, resources, includes*

Continuation of Table 4.2			
Topic	**Auto Label**	**Revised Label**	**Top 20-words**
6	**fire management**	**fire management**	*Indonesia, fire, forest, government, manage, haze, develop, cause, environment, nation, prevent, economy, intern, land, policy, cost, local , research, response, Indonesian*
7	**land cover**	**land cover**	*forest, land, plantation, log, cover, kalimantan, degraded, sumatra, swamp, palm, oil, loss, total, deforestation, use, distance, Indonesia, studies, map, change, concession, borneo*
8	**hotspot data**	**hotspot occurrence**	*hotspot, data, fire, land, cluster, forest, occurrence, spatial, province, studies, distribution, densities, use, distance, Indonesia, analysis, road, zone, location, detects*
9	**fire emission**	**fire emission**	*fire, emission, concentration, smoke, aerosol, air, atmosphere, haze, Indonesia, pollution, observes, burn, Singapore, region, studies, japan, Indonesian, biomass, forest, particle*
10	**charcoal site**	**charcoal site**	*record, site, date, charcoal, suggest, core, time, indices, human, reserve, increase, sample, acid, age, reduction, change, relate, accumulation, coastal, population*
11	**land use**	**livelihood**	*land, forest, cultivation, village, rubber, communities, local, farmer, agriculture, site, clear, rice, crop, people, studies, tree, plant, increase, product, burn*
12	**satellite data**	**satellite data**	*data, images, detection, vegetation, forest, burn, map, satellite, pixel, studies, modis, remote, sensing, cover, value, resolution, use, landsat, sensor, active*

Continuation of Table 4.2			
Topic	**Auto Label**	**Revised Label**	**Top 20-words**
13	**active fire**	**wildfire**	*fire, burn, fuel, ignite, active, vegetation, time, burnt, event, manage, cause, severe, occur, condition, spread, wildfire, includes, detect, source, scale*
14	**peat swamp forest**	**peat swamp forest**	*peat, peatland, water, carbon, swamp, ground, tropic, level, fire, surface, drainage, depth, table, emission, canal, kalimantan, central, dry, vegetation*
15	**soil moisture**	**soil moisture**	*forest, manage, species, soil, ecology, indices, nature, protected, degraded, agriculture, conserves, water, communities, biodiversity, habitat, product, secondaries, ecosystem, resources, includes*

4.4.3 Term interpretation

This section discusses the interpretation of terms from relevant topics as the contributing factors to the escalation of surface fire. Each topic from the LDA model assembles many specific common terms. But sometimes there are a few words that are common and highly probable in many other topics as well. As shown in Table 4.2, a few generic words occur in different topics. For example, term *forest* occurs in multiple topics, term *fire* is also found in more than five topics. Other terms that occur in multiple topics are *Indonesia* and *Kalimantan*. These terms occurred in many different topics because these words are used as the keywords in queries for the document collection. Term *data* also appeared in many topics such as Topic 1, Topic 4 and Topic 6. Term *model* occurs in two topics; Topic 1 and Topic 6. Other generic terms that occur quite frequently in various topics and have high probabilities in the word distribution is *studies*. These generic words do not represent the meaning of the topics. Therefore, the common and generic words are excluded in the interpretation process.

The purpose of this term interpretation is looking for the terms that can represent the contributing factors of peatland fires. Therefore, only topics that are related to the cause of forest and peatland fires were being observed. Even though a few subjects were already filtered out from the document collection, such as the impact of forest fire and policy-making, some topics that related to that research area still appeared. For example, the terms in Topic 4 and Topic 9 most likely represent the

Table 4.3: Topic and terms related to factors contributing to the escalation of peatland fire

Climate change	**Weather condition**	**Forest condition**	**Land cover**
change climate human agriculture	rainfall nino season drought enso temperature	degraded agriculture secondaries	land plantation cover degraded swamp palm oil
Hotspot occurrence	**Livelihood**	**Peat swamp forest**	**Soil moisture**
hotspot densities land use distance road	land cultivation village rubber communities farmer agriculture crop rice	water swamp level drainage depth table canal	soil content moisture depth

impact of forest fires such as carbon emission and smoke haze. Topic 6 seems to be a representation of fire management in Indonesia. These topics were excluded in the term interpretation. Other topics that were also excluded Topic 12 and Topic 13. Based on the terms inside each topic, the two topics represent the use of satellite imagery in the detection of a forest fire. These two topics were then excluded in the interpretation process. In addition to these two topics, Topic 2 was also eliminated. The elimination occurred because Topic 2 discussed the forest ecosystem, while the focus of this thesis is on the peatland area. The pattern in Topic 10 is indescribable, and even the generated auto label could not give a clear interpretation of the content of the topic. Therefore, this topic will not consider in the term interpretation process.

Table 4.3 shows the most likely terms for identification of factors contributing to the escalation of peatland fires.

Interpretation of terms - Climate change

The terms in this topic can be used to explain the effect of climate change on the forest condition in Southeast Asia. A few terms can be extracted from this topic and used to represent the contributing factors in peatland fire escalation. Terms *human* and *agriculture*, could be interpreted as human involvement in the occurrence of peatland fires. The way of people in Indonesia, especially Central Kalimantan, the land is prepared for agriculture by using the slash and burn method. This method is believed to be one of the causes of fire occurrence (Suyanto et al., 2009). The term *agriculture* could refer to the livelihood of people in an area. Therefore the term *agriculture* could be broadened into *livelihood*.

Interpretation of terms - Weather condition

The terms in this topic represent how the El Niño phenomenon influences the fire occurrence in Indonesia. A few terms can be extracted and used to represent the contributing factors. Term *rainfall* and *anomalies* can be used to explain the occurrence of fire in Indonesia that always happens during the dry season when the rainfall drops below normal. The fire occurrence is also triggered by the ENSO phenomenon, during El Niño conditions the amount of rainfall usually decreases significantly compared to non-El Niño conditions. The low amount of rainfall during the dry season in El Niño years also leads to drought conditions. The surface land, especially peatland, becomes dry and makes it easy to catch fire. Therefore, the terms *rainfall*, *ENSO*, *temperature*, and *drought* are chosen to represent the weather condition that influences fire escalation.

Interpretation of terms - Forest condition

In the topic of forest condition, three terms could be considered as contributing factors in the escalation of peatland fire, *degraded*, *agriculture*, *secondaries*. Fires are most likely found in the degraded land especially degraded peatland (Tacconi, 2003). Degraded peat has lost its capability to absorb water (Joosten et al., 2012), making this area easily dry out during the dry season and easily catch fire. Thus, the term *degraded* is included as one of the factors contributing to the escalation of peatland fires. The term *degraded* can be broadened into *land condition*, because the term *degraded* refers to the condition of land in one area. The term *secondaries* represents a type of the forest in which fires are most likely to happen. In Indonesia, especially Central Kalimantan, it is unlikely for fires to occur in the primary forests due to the closed forest canopies that make the forest always wet and humid (source). However, secondary forests are forests that experienced a significant natural and/or human-initiated disturbance of the original forest vegetation (Chokkalingam and De Jong, 2001). In this forest, fewer tree canopies are found. This makes the forest more prone to repeated fires (Harrison et al., 2009). Post agriculture activities are disturbances in the secondary forest that sometimes delay forest growth. The Indonesian government classified the secondary forest as one of the categories in land cover (, SNI). Therefore, instead of having *secondary forest* to represent a factor in the fire escalation, the term *land cover* was more considerable.

Interpretation of terms - Land cover

The terms occurring in this topic could be used to represent the land cover of an area that is affected by fire occurrence. Degraded terms as explained in the previous topic will be considered as contributing factors. In this topic, the terms *land* and *cover* most likely occur together referring to the phrase *land cover*. *Plantation* and *palm oil* are part of the classification of land use in Indonesia. Thus, the term *land use* was chosen instead of the extracted terms.

Interpretation of terms - Hotspot occurrence

The terms in this topic represent the relationship between land use and distance of road to the hotspot occurrence. As in Indonesia, the hotspot occurrence also can be used as representative of fire occurrence. Graham et al. (2014) reported that more hotspots are located in locations accessible by humans such as close to roads or rivers. Different types of land use also influences the hotspot densities. In the areas close to the settlement, fewer hotspots are found. Therefore, *road distance* and *land use* are considered as contributing factors. Term *hotspot* was also included, since it can be used as initial recognition of fire occurrence.

Interpretation of terms - Livelihood

In the topic of livelihood, a few terms can be used to represent the factors contributing to peatland fire escalation. The terms *plantation*, *rubber*, *farmer*, *agriculture*, *crop*, and *rice* represent what the local people in Indonesia, especially in Central Kalimantan, do for their living. Most of these activities involve the use of fires. When the villagers prepare one location for agriculture or farming, they used fire to clear the land. These fires sometimes escape and escalate into uncontrolled fires. Instead of taking all these terms as individual factors, all of them can be put under the term *livelihood*. Those terms could be considered as the states inside the contributing factor *livelihood*. Another term that could be interpreted as a factor contributing to the escalation of fire is *village*. Local people tend to clear and burn lands that are not too far away from their village. Therefore, the *distance of village* to the fire should be considered as a contributing factor.

Interpretation of terms - Peat swamp forest

There are some terms in the topic about peat swamp forests that could be considered as the contributing factors. *Water level* or *water table* refer to the phrases *ground water table* or *ground water level* that are mostly used in articles or reports about fires in peatland area. These factors refer to the explanation of how the changes in the water table influence the peatland fire (Usup et al., 2004). Since *ground water table* and *ground water level* have the same meaning, in this thesis the term *ground water level* is used. *Drainage* and *canal* also could be considered as contributing factors in the peatland fire escalation. Research found that construction of canals in the ex-Mega Rice Project had a devastating impact on the hydrology throughout the area (Dohong and Lilia, 2008). Through this canal's construction, the capability of the peat in retaining water had been impacted. Also ground water level in surrounding area of canals had lowered. These hydrology conditions have made the surrounding area of the canals prone to fires. Therefore, the terms *ground water level*, *canals*, and *drainage* are considered as the contributing factors. Since *canal* and *drainage* have the same meaning, the term *canal* was chosen.

Interpretation of terms - Soil moisture

A few terms in this topic could be considered as contributing factors in the peatland fire escalation. The term *soil* in this topic refers to peat soil. Thus, the terms *moisture* and *content* can be assumed as the phrase *peat moisture content*. The term *depth* refers to the depth of the peat soil.

As a result of the term interpretation process, 15 terms related to peatland fire escalation were gathered. These terms were then classified into three categories: 1) human activities; 2) climate conditions; and 3) biophysical factors and used as the variables/potential nodes in the structure of the causal model (see Table 4.5).

As shown in the literature review section 2.1.1, factors supporting peatland fire incidents could be classified into three categories: human activities, biophysical conditions, and climatic conditions. The terms that were extracted from the literature using topic modeling methods were fitted with these categories. The terms *rainfall*, *temperature*, *ENSO*, and *drought* can be used to represent the climatic conditions that support peatland fires and hotspots. The unique characteristics of peatland fires which are human-made was adequately represented by the phrases *road distance*, *canal distance*, *village distance*, and *livelihood*. The characteristic of peatland also can be captured from the existence of terms *ground water level*, *peat depth*, and *peat moisture level*. Through the terms extracted using topic modeling, we can produce terms or phrases that capture the contributing factors of peatland fire escalation.

4.5 Comparison Between Interpreted Terms from Topic Modeling and Terms from Experts

A set of terms was interpreted and selected as the factors that contribute to the escalation of surface peatland fire from the interpretation process in Section 4.4 (see Table 4.5). In this section, the terms from topic modeling were compared with a set of contributing factors provided by the fire experts. The purpose of this comparison is to measure how similar the result of topic modeling compares to human opinion and determine whether the result of topic modeling can be used as a starting point in the expert knowledge elicitation process.

4.5.1 Eliciting the Contributing Factors of Surface Peatland Fire Escalation Based on Expert Opinion

A group of experts was involved in the development of the causal model. These experts as explained in Section 3.1, Figure 3.1, have different knowledge backgrounds and expertise related to peatland fire in Indonesia. The experts were expected to provide a comprehensive list of contributing factors to represent the unique characteristics of peatland fire in Indonesia.

A focus group discussion was conducted to elicit the knowledge of the characteristic of peatland fire from experts in the development of the causal model, as explained in Chapter 3, Section 3.4.2. In this discussion, experts were asked about what they thought of the possible contributing factors in the peatland fire. Each expert's answers were written down and compiled to be a list of the potential contributing factors. From this process, 21 terms related to the contributing factors in peatland fires were elicited and compiled (see Table 4.4). However, there was disagreement between the experts on the list of terms that were presented. Some experts did not agree on some of terms that were used, because it could lead to confusion. An open discussion was conducted to revise the list. Each term then was analysed based on the correctness of the meaning and the familiarity of terms in the peatland fire research. Experts were asked whether each term already represented the meaning that it should. During the discussion, it was found that a few terms needed to change or be revised to make it more understandable. There are few terms that actually overlap with other experts and could be deleted.

Size of population is one term that could be misinterpreted because there might be two interpretations for this term. People can interpret this term as the number of individuals living in an area or the population density. The experts were saying that two factors have different influence in the peatland fires. If the focus is on the escalation of peatland, using the term *peatland population* will give a better explanation. Another term that could be merged is *vegetation type*. Since the term *land cover* also could be used to represent vegetation type in one location, it is concluded that there is no need to have vegetation type as another factor.

A few terms that have the same meaning were also discussed. Irrigation channels were mentioned by one of the experts as a contributing factor. However, waterways in hotspot distance from waterways represent both rivers and canals. In the ex MRP

Table 4.4: List of contributing factors provided by experts

1st version of terms	revised term
hotspot distance from waterways	hotspot distance from waterways
hotspot distance from settlements	hotspot distance from settlements
livelihood	livelihood
size of population	peatland population
surface fuel	fuel
land cover	land cover
land use	land use
vegetation type	
irrigation channel	
peat depth	peat depth
hotspot	hotspot
ENSO	ENSO
rainfall	rainfall
drought condition	drought condition
air temperature	air temperature
relative humidity	relative humidity
wind speed	wind speed
ground water level	ground water level
peat moisture content	peat moisture content
peat fires	fire escalation
canal berm	canal berm

area, irrigation channels are known as canals. Thus, the term *irrigation channel* could be merged with the term *hotspot distance from waterways*. The second column of Table 4.4 shows the revised terms that have been agreed on by all the experts in the group discussion. The revised terms then were used as the comparison list with the results from topic modeling.

4.5.2 Identifying the Similarity of Terms Resulting from Topic Modeling with Expert Opinion

The Jaccard similarity coefficient was used to identify the similarity between the list of terms extracted from literature and the list of terms provided by the experts. Based on the analysis using topic modeling, 15 terms were extracted. The experts provided 18 terms (see Table 4.5). The list of terms from topic modeling is labeled as Dataset A and the list of terms from experts is labelled as Dataset B.

The Jaccard equation from Equation 2.6 was implemented to calculate the Jaccard coefficient of the two datasets. The first step is finding the intersection of Dataset A with Dataset B, $(A \cap B)$. We found nine terms in Datasets A and B that were textually similar. The union of Dataset A and Dataset B generated 26 terms. The

Table 4.5: Terms classification

Interpretation	Topic model result	Expert result
Climate condition	*rainfall, temperature, ENSO, drought, hotspot*	*ENSO, rainfall, relative humidity, air temperature, wind speed, drought condition*
Human activities	*road distance, canal distance, village distance, livelihood*	*hotspot distance from settlement, hotspot distance from waterways, population density, livelihood*
Biophysical	*land cover, land use, ground water level, land condition, peat moisture content, peat depth*	*fuel, land use, land cover, peat moisture content, peat depth, ground water level, canal berm*

coefficient of similarity is determined by dividing the intersection result with the union of all terms in Dataset A and Dataset B.

$$SIM(A,B) = \frac{|A \cap B|}{|A \cup B|} = \frac{9}{26} = 0.35 \tag{4.1}$$

Based on the calculation, we found that the Jaccard similarity coefficient for the two datasets is 0.35 This value is closer to 0 instead of 1 which means that the two datasets are less likely to be similar. These dissimilarities are caused by: firstly, terms from the topic modeling that were not in the experts' list and terms from the experts that were not listed in the topic modeling; and secondly, terms with different names but similar meanings.

Jaccard similarity is used to determine the character–level similarity. It does not support the examination of meaning similarity. Based on further investigation on the dataset, a few of the terms were found to have the same meaning. For example, *temperature* and *air temperature*. Most the papers or research reports used the term *temperature* to describe the intensity of heat presence in an area. However, it can refer to the heat on the surface land and also air. Terms *village* and *settlement* also have the same meaning. They refer to the same meaning, a location where people live. Term *waterways* in the expert list refers to the existence of the closest rivers or canals, which is similar to *canals* in the topic model result. Therefore, a few revisions were made on the topic model result to adjust the terms used by the experts.

The Jaccard coefficient similarity was re-calculated using the revised dataset A, and the similarity coefficient increased to 0.62. This revised coefficient now is closer to 1. It means that the terms extracted from topic modeling are most likely similar to experts' opinions.

$$SIM(A, B) = \frac{A \cap B}{A \cup B} = \frac{13}{21} = 0.62 \tag{4.2}$$

Some dissimilarities were still found in these two datasets. Lestari et al. (2018) discovered that these dissimilarities were caused by terms that were never mentioned in the literature, such as *canal berm*; and terms that were rarely mentioned in the literature but are common knowledge in the peatland fire field, such as *relative humidity* and *wind speed*;

4.6 Chapter Summary

In this chapter, the first stage of the workflow for developing the causal model is presented. Identification of the causal variables, which are factors contributing to the escalation of surface peatland fires, was conducted using the implementation of LDA in the topic modeling method. This chapter also addresses Research Question 2.1 on how to automate information extraction from the literature to gather the contributing factors for the escalation of peatland fires. By implementing the LDA algorithm of topic modeling, many terms were extracted and could be considered as the contributing factors in the escalation of peatland fires. These terms can be classified into three categories which reflect the unique characteristics of peatland fire occurrences in Indonesia: climate conditions, human activities, and biophysical conditions. The terms are then compared to the list of contributing factors provided by the experts using the Jaccard similarity coefficient. The result from the comparison shows that the terms extracted from the literature using the topic modeling method are quite similar to the terms provided by the experts. This result indicates that in the situation where no domain experts are available to provide the causal variables for a causal model, automation of information extraction from the literature may be used to replace the domain experts' contribution.

These terms then are used as the initial talking points in the construction of the causal model for the second stage of developing the causal model. Further explanation about the development of the structure of the causal model using the initial terms is presented in the next chapter.

Chapter 5

Structure Development of a Causal Model

This chapter discusses the second stage of the workflow of causal model development that uses expert knowledge to construct the structure of the model. This chapter also addresses how to incorporate expert knowledge to develop a causal model that could be used to predict the escalation of peatland fire.

This chapter is divided into two parts. The first part contains an explanation of how to build and evaluate the structure of the causal model using knowledge from domain experts. The second part is the implementation of the second stage of structuring the causal model for surface peatland fire escalation.

5.1 Structuring the Causal Model

A causal model is used to explain the causal relationship between contributing factors of a phenomenon (Russo et al., 2011). In the literature review section 2.4, it was shown that the best way to present the relationship of each contributing factor is through graphical modelling, such as using Bayesian Networks (BNs) representation. Therefore, in this research, the BNs guideline is followed to develop the structure of the causal model. The identification of initial variables, known as nodes, has been done in Stage 1 of the workflow. These nodes are then used as the initial talking points in the development of the causal model structure in Stage 2 of the workflow.

Structuring the causal model in the proposed workflow (see Figure 5.1) starts with establishing the objective or goal of the causal model, followed by presenting the initial nodes to the experts and asking the experts to create the relationship of each node. Since there is more than one expert involved, the experts' answers need to be combined. The structure of the causal model is presented in an open discussion. This structure is then evaluated and refined, as needed, before adding the states for each node. Evaluation and refinement of the causal model structure is an iterative process until the structure is deemed acceptable by the domain experts.

In the next subsection, a detailed explanation of how to build the structure of the causal model is presented.

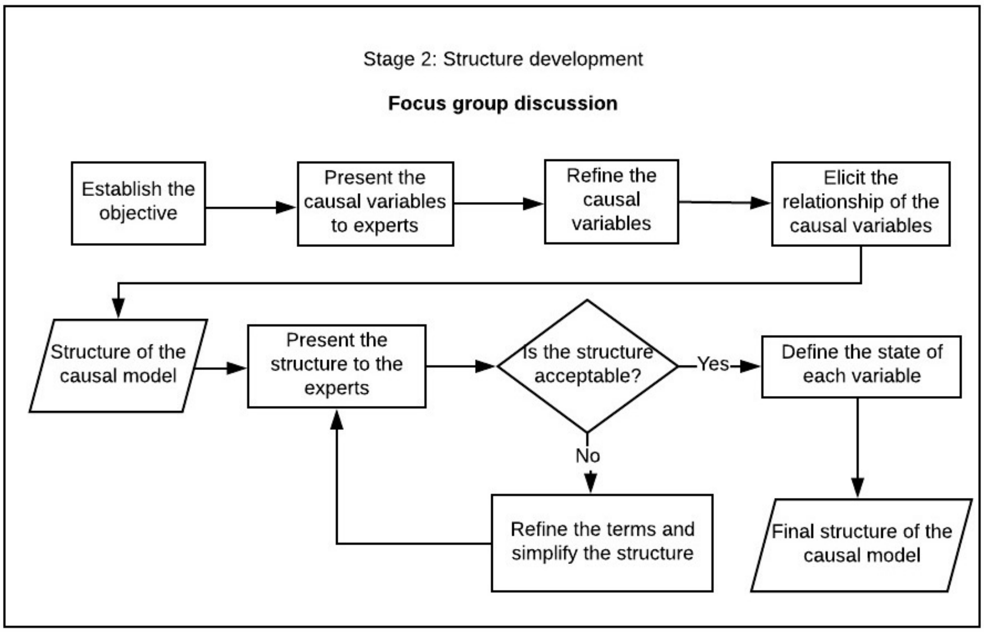


Figure 5.1: Steps in the Stage 2-structure development of a causal model

5.1.1 Establishing the Objective for the Causal Model

Before creating the structure of the causal model, the objective of the causal model is presented to the experts. The experts are then asked to comment on this objective as it relates to their expertise and experience. Cain (2001) suggested a few questions that can be used as the guideline:

- do you [expert's name] think that achieving these objectives is important?
- what other objectives do you [expert's name] think are more important?

The responses from the experts to these questions could influence the objective. It can widen the original objective or narrow it down. The criteria of the objective needs to be defined by asking the experts the ways in which this objective could be achieved. Once all the experts agree on the objective and its definition, the next step is to identify the node for the causal model.

5.1.2 Refine the Causal Variables/Nodes

The initial variables/nodes were identified through the automation literature review and experts' opinions as covered in Chapter 4. The nodes then are presented to the experts in an open discussion. The way of presenting the variables to the experts can be done in various ways. Examples are sketching the variables on a whiteboard, printing out the variables on a plain paper for each expert, or displaying using other devices. Using any tools that enable the easy drawing or redrawing of the variables helps the elicitation process flow (Mendes, 2014). The description of each variable is then explained to the experts and the meaning of each variable is bound by what they represent within the context of a problem domain. To gather the same understanding on one variable is not an easy task as experts have a different level of knowledge and expertise. Taking one step at a time is the best way to facilitate a fruitful elicitation exercise. Using a tangible and real scenario related to the problem domain will keep the experts focus on the context of the variable definitions. Focusing on one of the variables at a time, and getting each of the domain experts to speak and voice opinion on the meaning of the variables. It is critical to always ask experts to justify their opinions. Asking "why" questions is also necessary so the experts can ground their suggestions.

Once all the experts agree on the nodes, experts can be asked to determine the relationship of each node. The process of creating the structure is presented in the next subsection.

5.1.3 Creating the Structure of the Causal Model

If the experts involved in the elicitation process are not familiar with the causal relationship, it is necessary to start the process with a simple explanation of what is meant by a cause and effect relationship. A simple model that shows the causal effect relationship of forest fire is presented in Figure 5.2. This simple model shows that there is a causal link between human involvement and forest fire, and climatic condition and forest fire. Similar relations exist for possibilities of smoke occurring as result of forest fire, and land degradation occurring due to the fires. Through this simple example, it is expected that experts could have a better understanding of the notion of causal effects and could relate this to the problem domain.

Experts were consulted and asked whether the presence of one node influences other nodes. The question could be in the form of a **Yes/No** question, such as *does node X influence node Y?* to reveal if there is a direct relationship between two nodes. The causal model contains a set of nodes. To obtain the relationship of each variable,

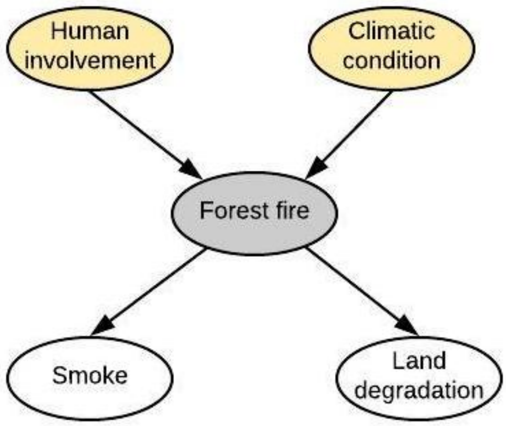


Figure 5.2: A simple illustration of cause and effect relationships in the forest fire problem

the questions should cover all combinations of two nodes. An n x n adjacency matrix can be used to generate the possible questions for the nodes. This adjacency matrix contains all possible direct cause-effect relations among the nodes (Nadkarni and Shenoy, 2001). Based on the example of a simple model in Figure 5.2, the directed arcs could be encoded as entries in the adjacency matrix below in Table 5.1 below.

For each pair of nodes, experts are asked to specify if there is a directly connected causal relation. A direct arc is used to indicate the direction of the effect.

Table 5.1: Example of an adjacency matrix for forest fire problems

	Human involvement	Climatic condition	Forest fire	Smoke	Land degradation
Human involvement					
Climatic condition					
Forest fire	✓	✓			
Smoke			✓		
Land degradation			✓		

5.1.4 Collating Answers

The previous subsection discussed how the expert knowledge could be used to construct the structure of the causal model. Most likely for a complex and varied domain problem more than one expert is involved in the development of a causal model (Cain, 2001; Nadkarni and Nah, 2003; Pollino et al., 2007). If there are multiple experts involved, multiple responses will be obtained and the structure of causal model is developed based on these multiple responses. Therefore, the next step in this stage is collating multiple responses from experts to a single structure of the causal model.

There are few methods for collating expert answers. A majority vote algorithm is one of the simplest methods in the implementation of collating multiple responses (Lam and Suen, 1997). When using the majority voting to combine experts' answers, no prior knowledge about the participants is needed. It also does not require training on a large dataset. The number of experts answering **Yes** to a question of the form *does variable X influence variable Y?* is calculated. If the majority of experts says it does or **Yes**, then a direct arc between X and Y is inserted into the causal model structure. Otherwise no direct arc will be put between X and Y.

5.1.5 Evaluating the Structure of a Causal Model

Once a single structure of a causal model is created, it is time to use any available BNs tool to present all the variables and their relationship together with the states. However, before implementing the structure of the causal model with a BNs tool, it is necessary to ensure that the structure of causal model follows the standard guidelines of BNs as described in the literature review. The experts also need to agree on the structure of causal model. Therefore the next step is evaluation and refinement of the causal model structure.

The evaluation of the structure is conducted as an iterative process. It is aimed at evaluating the nodes and their relationships until the structure of the causal model is acceptable by the domain experts. This evaluation process covers discovering whether an important arc is missing or that an arc should not be there, solving the variables forming circular links, restructuring the causal model based on compactness and node-ordering rules, and generating the states for each of the variables.

Feedback Loop. The structure of a BN is a directed acyclic graph (DAG). Thus, it is not possible to have cyclical loops such as feedback loops within a BN (Jensen and Nielsen, 2007), although feedback can exist between two nodes to show dynamic changes over time. Therefore, Korb and Nicholson (2011) suggested modelling the feedback process with a dynamic Bayesian Network (DBN). However, Pollino and Henderson (2010) stated that if a model does not require changes to be presented dynamically, additional nodes can be added to represent changes over time. An intermediate node can be created between the two nodes to solve the feedback loop problem.

Compactness. If a node has too many parents, then the conditional probability table (CPT) can become very large. One way to reduce the number of parents is by 'divorcing' the multiple parents (Boneh et al., 2006). Some of the parents of a node could be removed or divorced by introducing a new intermediate node that summarises the effects of a subset of parents of a child (Olesen et al., 1989).

Childless node. If there are some nodes in the structure which have no children, these nodes should be investigated. If these nodes describe the objective of the causal model, then the nodes should be included in the structure. However, if the childless nodes cannot be used to describe the objectives then the existence of these nodes in the structure should be examined (Cain, 2001). If a direct link could be connected from this childless node to the objective node or an indirect link through other nodes, this node could be kept in the structure. However, if no direct or indirect link could be made between the childless node and the objective node, then this node are probably unnecessary and could be removed.

Direction of the arcs. A set of questions is used as guidance in the structure evaluation and discussed with the experts. The question *are the directions of the arcs right?* was also used to evaluate the structure. The structure of the model should be changed if the experts answer "no" to any of the questions.

Clarity of the variable names. All the variables and their values should have a clear operational meaning (Mendes, 2014). A question such as *are the nodes in the structure the right ones? are they named properly?* could be used in this evaluation.

D-Separation. Jensen and Nielsen (2007) defines if two nodes X and Y in the causal network are d-separated for all links between X and Y, there is an intermediate node Z such that either:

- the connection is serial or diverging and Z is instantiated, or
- the connection is converging, and neither Z or any of Z's descendants have received evidence.

5.1.6 Defining the State of Each Variable

The next step is to determine how each variable will be measured. This includes how many categories or states each variable should have and which states are to be used to measure each of the variable. The elicitation process can be started by asking the experts such questions (Korb and Nicholson, 2011): *what values can the variables take*, or *what state can they be in*? The decision of these states will affect the number of probabilities to elicit, therefore (Marcot et al., 2001) suggested limiting the number of states to five. However, this limitation should not constrain the experts in providing and deciding the states that should be used. The number of states really depends on what is considered most important in the problem domains (Mendes, 2014). The states should be mutually exclusive and exhaustive. They can take discrete values such as Boolean (true or false), ordered values (low, medium, high) or integral values (the ranges of values which the variable can take). If there is a recognised classification or management guideline available, the state also could be generated from these sources.

Once all the experts agree with the variables and the states, the next step is to elicit the probability.

5.2 A Causal Model for Predicting Hotspot Escalation in Peatland Fire Central Kalimantan, Indonesia

The process of development of the causal model structure, described in the previous section, was then implemented to build a causal model for peatland fire escalation in Central Kalimantan, Indonesia. The purpose of constructing this causal model is to describe and explain the characteristic of peatland fire escalation in Indonesia. By having an comprehensive understanding of why the escalation of peatland fire happens, a predictive model of future fire as a prevention tool could be developed.

In this process of developing a peatland fire causal model, fire experts were invited: five from Indonesia, one from the USA, and one from Australia. These experts were chosen based on their expertise and practical knowledge of Indonesian fires. The fire experts from Indonesia were involved in order to gather information based on their local wisdom and real-life experience of peatland fires. The fire expert from the USA was invited to gather more knowledge in the human dimension of peatland fires. The fire expert from Australia was invited to obtain information about climatic conditions related to forest fires.

These experts were involved in the focus group discussions and each expert is labeled with E1 - E7. The explanation about the profile of each expert is provided in Appendix D.

5.2.1 Determining the Objective for the Causal Model

In this research project, establishing the objective of the causal model is one of the *problematic* processes. It took 1.5 days for the experts to reach consensus on an objective. The original objective of the causal model is to predict peat fire occurrence based on hotspot information. However, a few experts found this objective was not really important and proposed another objective. They pointed out the importance of exploring forest fires in general rather than focus on peat fires. However, other experts argued that in Central Kalimantan, fires happened mostly on degraded peatland areas. In a degraded area, no forests exist anymore. Therefore fires are known as peatland fires. Therefore, the idea of changing the objective to forest fire was ignored.

At first, experts pointed out that this objective was too broad. As is classified in Usup et al. (2004), peat fire is categorised into surface peat fire and deep peat fire. This issue was raised by two experts:

> " ...for peat fire there is deep peat fire or surface peat fire? So what kind of fire that you want to model?"(E3, 13/02/2017, min:02:06:20).

> " ...the terminology that used to use in Central Kalimantan is peatland fire. ...it happens on the surface of the peat"(E2, 13/02/2017, min:02:06:20).

The objective then was to narrow down to the surface peatland fire. Once again, experts were asked whether this objective was relevant to the domain area or needed to be refined. Some experts suggested specifying an objective based on what should be investigated on a surface peatland fire. One expert mentioned, based on the experience on the ground, that it is difficult to decide whether a fire will escalate based on hotspot information. If there is information on which hotspot might escalate, it will help to decide which area needs more attention or preventive actions.

Some experts suggested a few definitions of escalation:

> " if there is a fire hotspot occur in one day and in the next day there are a couple hotspot occur nearby, the it counts as escalation."(E4, 14/02/2017, min:02:06:20).

> "the escalation is if there is a huge fire happen in the next day"(E1, 14/02/2017, min:02:06:20).

> "if there is a spot fire [identify from the hotspot occurrence] ...then are multiple of hotspot occurred within the certain radius (maybe 2km, 3km) from the original hotspot in next day or 2 days or 3 days"(E2, 14/02/2017, min:02:06:20).

The experts then agreed to conclude the definition of escalation as the fires detected beyond the initial hotspot within the area of 2 by 2 km^2 (within a period of 1-2 days).

5.2.2 Refinement of Causal Variables

At the beginning of the focus group discussion, 19 initial variables as the result of Stage 1 in Chapter 4 were presented. All of the variables were sketched out using an oval shape (see Figure 5.3) and displayed it to the experts. The experts were asked whether any important variables had been overlooked or if irrelevant variables had been included.

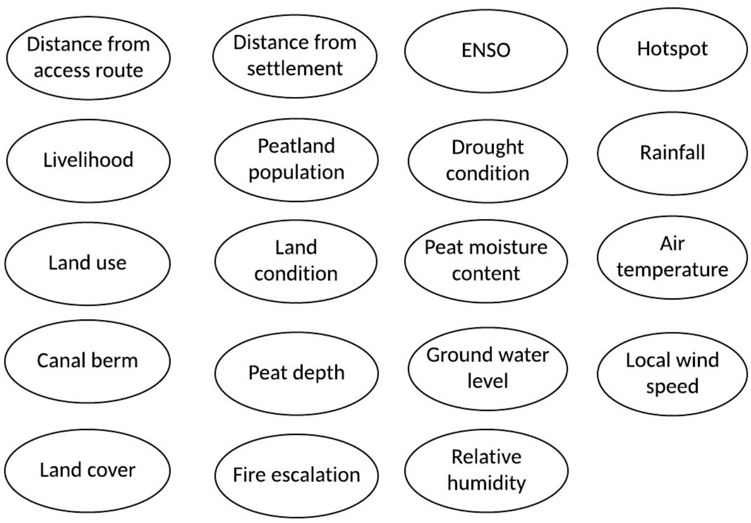


Figure 5.3: Variables for a surface peatland fire escalation causal model

Land use and land cover change. There are two variables that consider overlapping, *land use* and *land cover*. Experts suggested combining these two variables and setting the name to *land use land cover* (see Figure 5.5).

5.2.3 Creating the Structure of the Causal Model

Once the experts agreed on the causal variables or nodes, the next step was determining the relationship of one node with the other nodes. A sketch of all variables were given to each experts and they were asked to draw the relationships of one variable to the other variables. A question on whether the presence of one variable influences other variables: *Does variable X influence variable Y?* was used as guideline to draw the links/arcs on each node.

Experts were given 30 minutes to 1 hour to draw the relationship of variables on sketch paper. The experts had been able to provide a causal relationship from each node. Figure 5.4 shows one of the expert responses. At this point, there is no examination whether the causal relationship follows the BNs guideline or not.

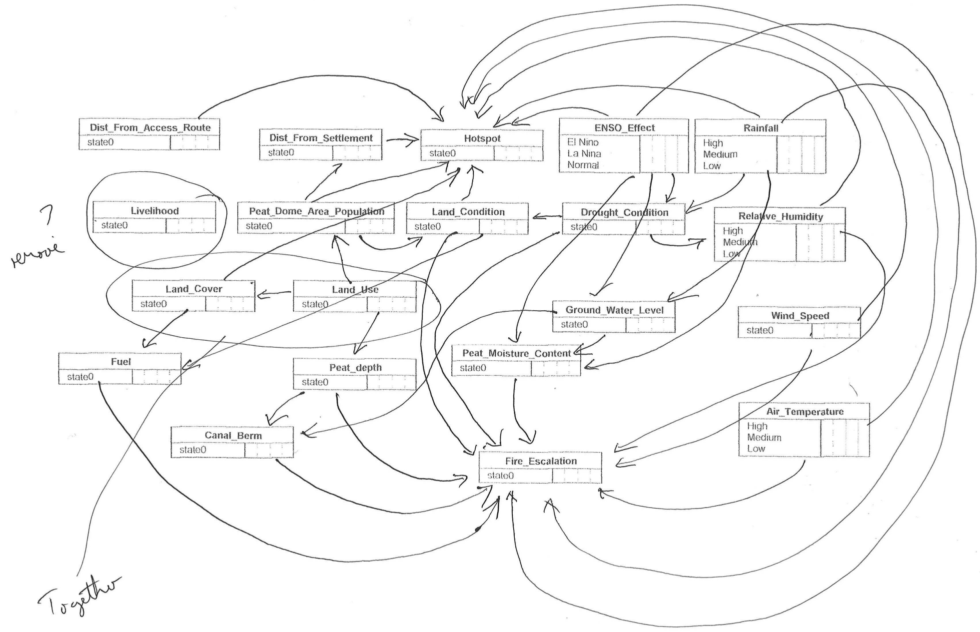


Figure 5.4: Sample of the expert's sketch on the relationship between causal variables

All expert responses were combined using majority voting. Table 5.2 shows the adjacency matrix based on the combination of expert responses. If the majority of experts drew a direct arc between two nodes, then it was encoded in the matrix. Based on this matrix, a single structure of peatland fire escalation is generated (see Figure 5.5).

5.2.4 Evaluation of the Causal Model Structure

The structure of the causal model that was collated from the experts' answers was complicated and messy (see Figure 5.5). This structure containing a child node with too many parents, for example child node *fireescalation* has 11 parent nodes that connected to the child node *fireescalation*. The structure also has a couple of relationships with feedback loops. The structure needed to be refined and evaluated to achieve the most straightforward structure. As mentioned in Subsection 5.1.5, this evaluation process is an iterative process. This research project conducted two

Table 5.2: Adjacency matrix of majority voting results for the surface peatland fire escalation causal model

Effect \ Caused	Dist.from access route	Dist.from settlement	Livelihood	Population Density	Land use land cover	Canal berm	Fuel	Land condition	ENSO effect	Rainfall	Air temperature	Relative humidity	Local wind speed	Ground water level	Peat moisture content	Fire escalation
Dist.from access route				✓												
Dist.from settlement			✓	✓												
Livelihood				✓	✓											
Peatland Population			✓				✓									
Land use land cover							✓	✓								✓
Canal berm							✓									✓
Fuel																✓
Land condition					✓											
ENSO effect							✓				✓					
Rainfall								✓				✓		✓		
Air temperature																✓
Relative humidity																
Local wind speed																✓
Ground water level																✓
Peat moisture content																✓
Fire escalation																

iterations of evaluation in one focus group discussion. The first iteration was aimed at solving some problematic aspects of relationships such as feedback loops and relationship compactness. After the first iteration, the structure of the causal model was presented again to the domain experts. The second iteration of evaluation was conducted to examine some of the nodes and whether they were still relevant with the revised structure.

In this first iteration, the structure of surface peatland fire escalation was presented to the domain experts (see Figure 5.5). An open discussion was then conducted for the experts to collaborate on refinement and evaluation of the causal model structure based on the guidelines provided in Subsection 5.1.5

5.2.4.1 Solving the Feedback Loop Problem

In the structure of causal model presented in Figure 5.5, two feedback loops were identified. First is the relationship between livelihood and peatland population; second is the relationship between land condition and land use. As mentioned in

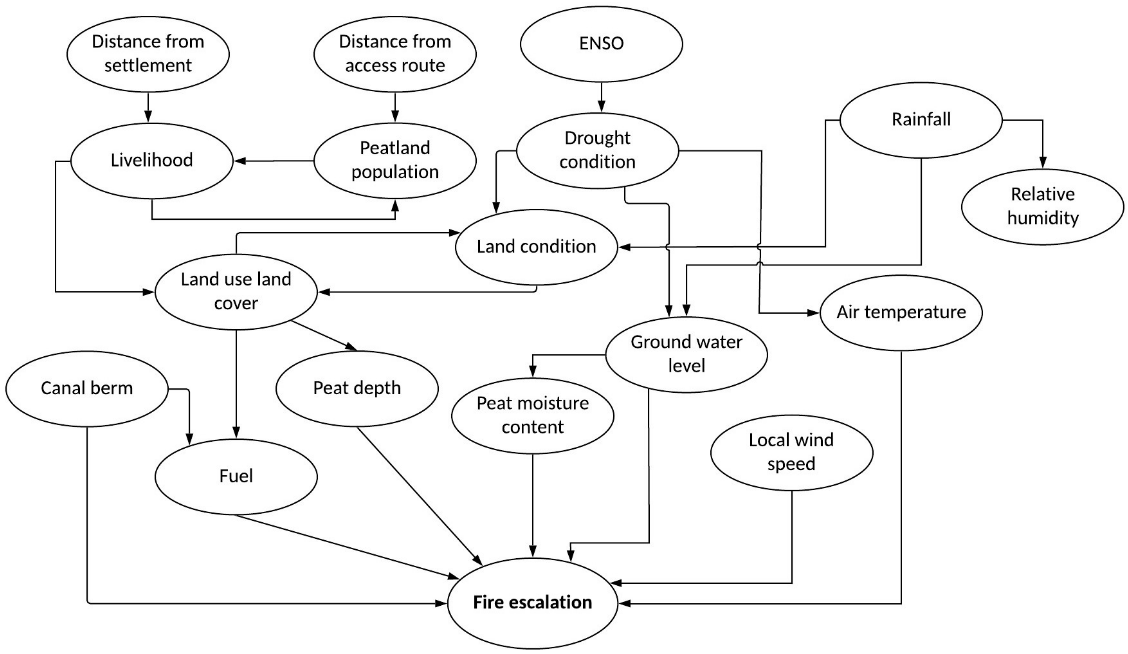


Figure 5.5: Version-1 on the structure of the surface peatland fire escalation causal model. This structure is the result of the 1^{st} iteration in structure development, created based on the survey from the fire experts.

Subsection, the feedback loop can be solved using DBN or adding an intermediate node. Since the changes in this causal model do not need to be presented dynamically, adding a new intermediate node was chosen to solve the feedback loops.

The first loop was between node *livelihood* and *peatland population*. The majority of experts drew a relationship between these two nodes because both of them affected each other. Two experts explain the connection using different examples:

> "... because the population will affect also the style [livelihood] of people in this area, that is why there is link [from population] to livelihood then."(E1, 14/02/2017, min:04:07:48)

> "... the attractions of livelihood... [makes people come to the area] Gold mining maybe is the example... which attract people from other area to come"(E3, 14/02/2017, min:04:08:28)

In this case, livelihood or lifestyle in an area are most likely to attract new people to move and live in that place. For example, the existence of gold mining in an area could attract people from different areas, even regions, to move there and join the gold mining. The peatland population also determines the livelihood of most people in the area. This assumption has led to creation of feedback loops between nodes: livelihood and peatland population (see 5.6(a)). Since the livelihood and peatland population affect each other at the same timestamp, changing the structure into a DBN is not an option. Instead, one expert suggested adding a new node between *livelihood* and *peatland population*:

> " ... you should find something [new node] in the middle between livelihood and peatland population. So the you can draw here[livelihood] from here[new node] and from here [new node] to here[peatland population]."(E1, 14/02/2017, min:04:11:47).

> " ...the livelihood affects something thing[new node] that we will have new term then the peatland [population] also affects this new term[new node], this new term affects the land use land cover[node] change..."(E1, 14/02/2017, min:04:14:00).

This new node becomes an intermediate node between livelihood and peatland population. The relationship of this intermediate node with livelihood and peatland population refers to economic activities. van Beukering et al. (2008) stated that in the peatland area, the economic activities are dependent on how local communities use the natural resources, especially logging or agriculture. Since the livelihood or attraction in the peatland area could attract people to come, which might then increase the population. Some experts mentioned that the number of people living in one area affects the economic activities. Therefore, experts agreed to name the intermediate node with *economic activities*. Having this intermediate node (see Figure 5.6(b)) has resolved the loop problem without losing the causal relationship between the nodes.

> "the economic activities/attraction happening in one area attract people to come and make fire "(E1, 14/02/2017, min:04:14:00).

> "population, livelihood ...., all these should affect the economic activity "(E2, 15/02/2017, min:01:10:35).

The second loop was between node *land condition* and *land use land cover*. The majority of experts noted that the condition of land, whether it is degraded or not degraded, will change the land use and land cover in one area. For example, degraded land most likely will not have primary forest anymore as land cover. This condition will change the land use land cover type on that area from primary forest to open area. The land use and land cover type on an area also will influence whether the condition of land becomes degraded or not. To solve this feedback loop problem, one

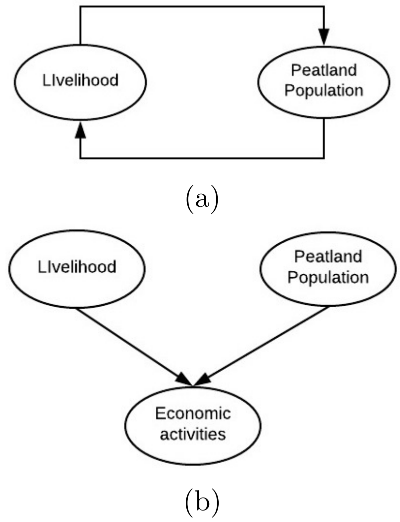


Figure 5.6: Feedback loop between livelihood and peatland population. (a) nodes with feedback loop; (b) nodes without feedback loop

expert once again suggested having an intermediate node. As mentioned in one of the quotes above, the intermediate node between livelihood and peatland population could be used as the intermediate node for the feedback loop in land use land cover and land condition (see Figure 5.7). The level of economic activities of people in one area could influence the changes of land use land cover and also land condition. The more economic activities done in one area most likely will change the land use land cover and the condition of land. The experts also recommended keeping the relationship from the *land condition* node to the *land use land cover* node, since the combination of economic activities and land condition could influence more on the land use land cover.

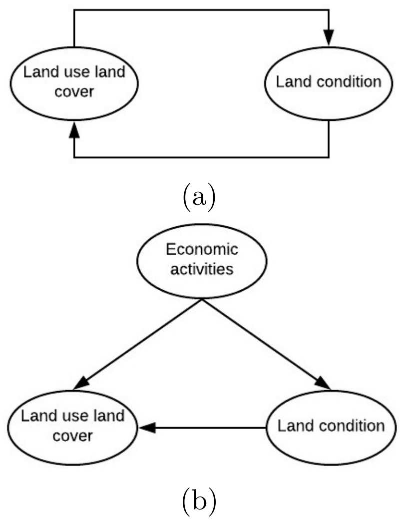


Figure 5.7: Feedback loop between land use land cover and land condition. (a) nodes with feedback loop; (b) nodes without feedback loop

5.2.4.2 Compactness of the Multiple Parent Nodes

As in Figure 5.5, node fire escalation has seven parents. This condition could cause a problem in the parameterisation process, because in order to create the CPT for that node, a large number of questions will be generated (Friedman and Goldszmidt, 1998; Zagorecki et al., 2006). One way to resolve this problem is to divorce these multiple parents and add a few intermediate nodes.

The first intermediate node aimed at grouping the nodes is related to climatic conditions. Nodes for rainfall, relative humidity, wind speed, and air temperature were grouped into one intermediate node *degree of fire danger* (see Figure 5.8(a)). By adding this intermediate node, two parents for the child node peatland fire escalation can be represented by only one parent. One expert suggested how this could be done:

> "Those rainfall, wind speed, relative humidity and air temperature are all the inputs for fire danger index ... and then one arrow coming to fire danger index to surface fire [node]"(E4, 14/02/2017, min:04:17:47).

A new intermediate node was also added to represent the peat condition as seen in Figure 5.8(b). This node, named *peat flammability* was used to group nodes for ground water level and peat moisture content, and aimed to show the flammability of one area based on the dryness of the peat. The dryness of the peat could be measured from the moisture content that was influenced by the ground water level. A new node was introduced as the parent of an intermediate node. Peat flammability was not only influenced by the moisture content but also by the decomposition level of peat in that area. The low moisture content and dry weakly decomposed peat makes an area highly flammable (Usup et al., 2004). Therefore, node *peat decomposition level* was added to the structure and it has a direct arc to the *peat flammability* node.

Reducing the number of parents also can be done by restructuring the arc directions. In Figure 5.5, there are three nodes that could be grouped into one node, which was used as a parent to the fire escalation node. *Land use land cover* and *canal berm* could represent the fuel condition. Therefore, the direct relationship of these two nodes to fire escalation should be changed to *fuel flammability* node as shown in Figure 5.8(c) .

5.2.4.3 Removing Irrelevant Nodes

There is one node, *hotspot*, that did not have any links to other nodes. In some of the sketches collected from the experts, they indicated that this node should be removed.

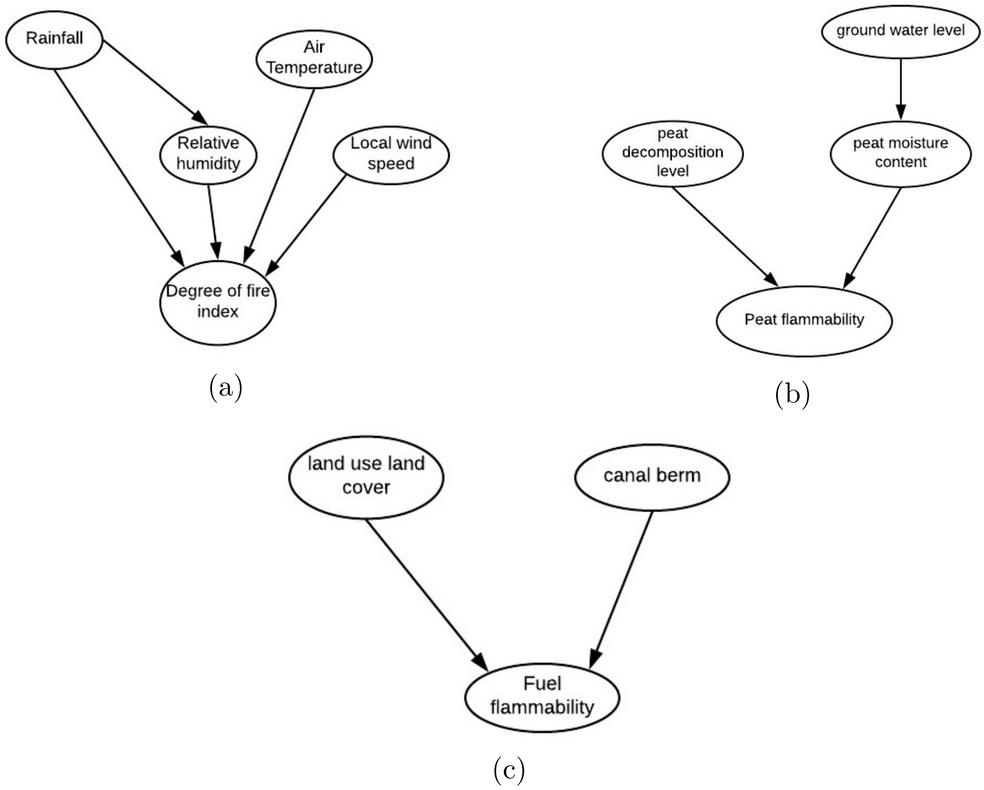


Figure 5.8: New intermediate nodes reduce the number of parents of the fire escalation node. (a) degree of fire danger (b) peat flammability (c) fuel flammability

One expert argued that a hotspot is only used as an indicator of fire occurrence. The expert suggested hotspots might escalate into wildfires, but they are not the contributing factors in the escalation of fire in peatland areas.

> "Hotspot is only the indicator there is increasing temperature on the ground. .... it is not always true, when the number of hotspot reduce or not hotspot at all it means no fire."(E1, 13/02/2017, min:04:07:48)

Peat depth is another node that seems irrelevant, as the objective of this causal model is the detection of the surface peatland fire escalation. The depth of peat should not influence the escalation of the fire on the surface. One expert mentioned that the depth of peat such as 0.5m or 1m will not affect the escalation:

> "As we talking about escalation on the peatland, whether it is half metre or one metre of peat. I do not think it affect the escalation."(E2, 14/02/2017, min:04:07:48)

A new structure of the causal model was created based the evaluation above, as shown in Figure 5.9. The next step was to determine the states from each node. However, in the process of defining the states, the structure of causal model was evaluated again in order to get better reasoning with the states. Subsection 5.2.5 presented the second iteration on evaluating the structure of the causal model.

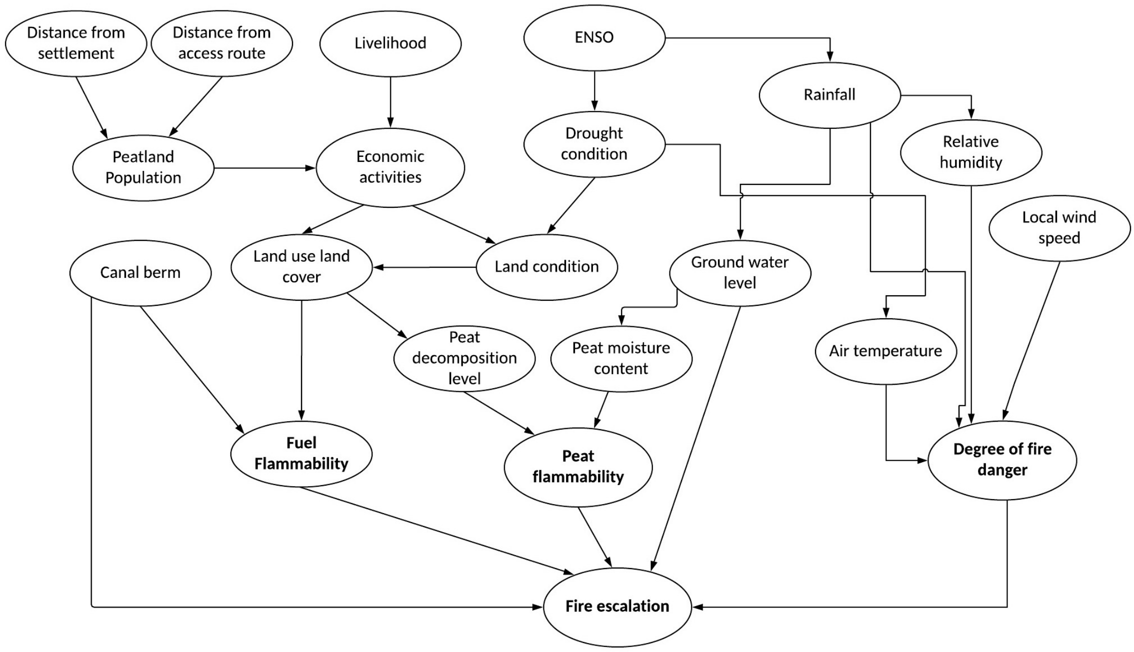


Figure 5.9: Version-2 on structure of the surface peatland fire escalation causal model after revision. This structure is a result of the 2^{nd} iteration in the structure development, revised based on the input from the experts.

5.2.5 Refinement of the Causal Model Structure - Second Iteration

The first iteration of the causal model created a new structure for the peatland fire escalation causal model as shown in Figure 5.9. The sketch of the revised structure was presented again to the experts. In this second iteration, experts were asked if there were still any irrelevant nodes or overlooked nodes. This iteration also covered the clarity of the node descriptions and their relationships together with adding states to the nodes.

5.2.5.1 Removing Irrelevant Nodes and Links

Even though this is the second version of the causal model, there is still a possibility of overlooked nodes or irrelevant nodes that need to be examined. Therefore, in this iteration experts still found some of the nodes that could be excluded from the structure.

Relative humidity and ***air temperature*** were the two nodes that suggested to be removed from the structure. One expert argued the changes of values in these nodes would not affect the degree of fire danger:

> ".. relative humidity [in Central Kalimantan] is always high, air temperature is always warm to high... I think if we want to make it[the model] simpler we could take out those variables and just use rainfall "(E4, 15/02/2017, min:01:54:47).

However, two experts have different opinions on why relative humidity should be kept in the structure:

> "Even though in general the relative humidity in Central Kalimantan is high (about 80 – 90%) but during dry season (when fire happened, in October), the humidity drops into 60% "(E2, 15/02/2017, min:01:58:42).

> "We cannot delete all the variables (relative humidity, air temperature) that influence degree of fire danger, because they are part of the fire danger index calculation. "(E1, 15/02/2017, min:02:06:20).

Therefore at the end, experts agreed to keep the relative humidity in the structure due to the possibility of variation during the fire season, and removed the air temperature from the structure. Since the air temperature was deleted, the relationship from drought conditions to air temperature was also removed. Thus, the parents of the degree of fire danger node were now reduced to only three.

Drought condition was also one of the variables suggested for removal. The meteorological condition of drought is a condition when there is a prolonged period of rainfall being less the average, causing a deficit of available water (Schweithelm and Glover, 1999). In Indonesia, the low pattern of rainfall may lead to the drought condition (Vernimmen et al., 2012). This is supported by one argument from an expert about why it is unnecessary to have a drought condition as it is already represented by rainfall:

> "I am not sure about what [information] we will get from drought condition that we will not get from the rainfall."(E2, 15/02/2017, min:02:00:00).

Therefore, the drought condition node was removed and the condition in rainfall nodes is used to represent the drought condition.

Ground water level. In this examination, some experts drew attention to an unnecessary link between *ground water level* and *fire escalation*. As shown in the Figure 5.9, there was a direct link from the *ground water level* node to the *fire escalation* node. *Ground water level* also has a direct link to *peat moisture content*, while *peat moisture content* has a direct link to *peat flammability* which influences *fire escalation*. Two experts found these relationships were redundant. They suggested keeping only *ground water level*, removed *peat moisture content* and change the direct link of *ground water level* to *peat flammability*.

> "We can delete peat moisture content since we have ground water level contribute to the surface fire escalation and peat moisture content ... just keep ground water level."(E2, 15/02/2017, min:01:40:17).

> "We can have direct relationship from ground water level to the peat flammability and peat flammability to fire escalation ... The peat flammability explanation includes the peat moisture content"(E1, 15/02/2017, min:01:41:00).

However, not all experts agreed with these arguments. One expert insisted on keeping the ground water level and changing the link to peat flammability through peat moisture content:

> "Peat moisture content is important and should be included in the model [structure]. Ground water level is affecting peat flammability through peat moisture content." (E5, 15/02/2017, min:01:40:29).

At the end, the experts agreed to simplify the relationship of ground water level to the goal node, *fire escalation* node (see Figure 5.10). This node then only has a direct link to peat moisture content, since the level of ground water influences the amount of water or the moistness of the peat. If the peat moisture content is low then the peat become more flammable.

5.2.5.2 Refining the Order of the Nodes and the Links

It is essential to make sure the structure of the model shows the cause – effect relationship (Korb and Nicholson, 2011). However, sometimes experts make mistakes with the order of the nodes and the direction of the arc. For example, some experts suggested that the cause of changes in population density in one area was the distance of fire from the settlement and access route. However, after further discussion we found that this cause–effect should be flipped, as shown in Figure 5.11.

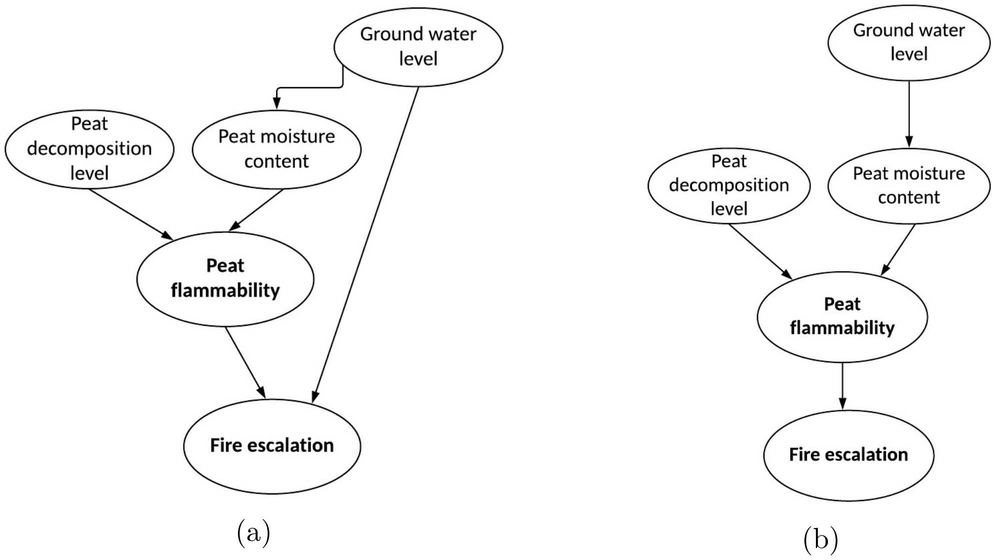


Figure 5.10: Feedback loop between land use land cover and land condition. (a) nodes with feedback loop; (b) nodes without feedback loop

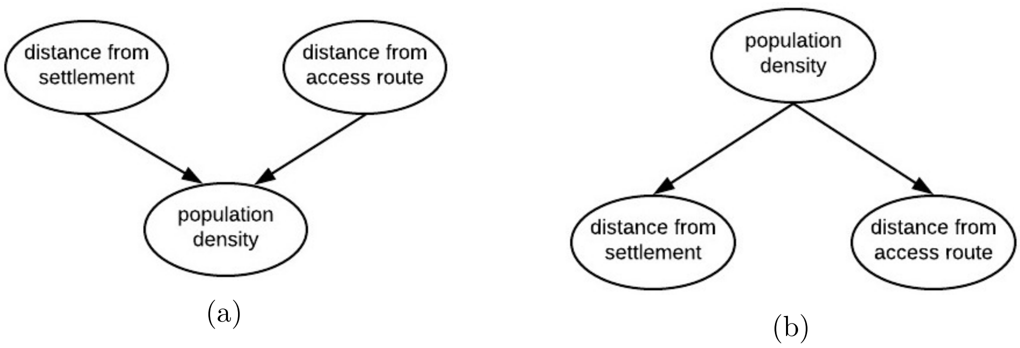


Figure 5.11: Feedback loop between land use land cover and land condition. (a) nodes with wrong directions of arcs; (b) nodes with correct directions of arcs

After these nodes were flipped, the node *distance from a settlement* and node *distance from an access route* became leaf nodes or nodes without children. This revealed another problem, because it meant these two nodes had no influence on any of the nodes related to peatland fire escalation. But, it is known that the escalation fire in the peatland area is most likely influenced by human activities. The availability of human access to an area was one of the factors that represented human activities. As mentioned in 5.1.5, the childlessness problem is solved by adding a direct or indirect link to the objective node, *fire escalation*. An indirect link was created by adding a new intermediate node to represent the influence of human access to the fire escalation. This intermediate node was labeled *fire susceptibility* and represents the vulnerability of an area to a fire regarding accessibility of that area. The node distance from a settlement and the distance from an access route became the parent nodes of *fire susceptibility*. A direct link to the objective node is created from this intermediate node. This representation of the new intermediate node is presented in Figure 5.12

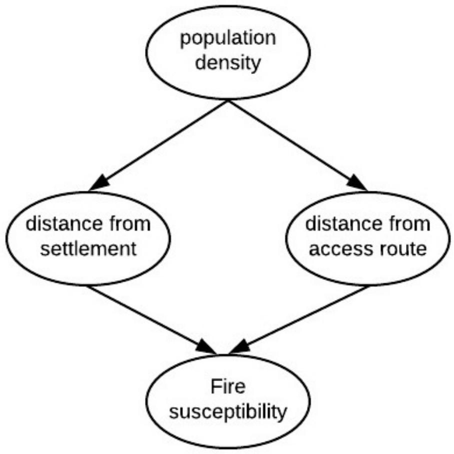


Figure 5.12: Adding fire susceptibility node

5.2.5.3 Clarity of the Node Name

In this second iteration, each node was examined again to ensure it had a clear operational meaning.

Fire escalation. Experts argued that the name for this node should be clearer. Since the focus is on the peatland area and surface fire, the name of this variable should be changed to *surface peatland fire escalation*. The description of this node was also redefined based on the time and area of escalation.

Peatland population. The label of this node was still a bit confusing for some of the experts. For example, one expert asked about the unit for this node:

> “What is the unit of this [population]? .... is it one village or the whole area? ”(E3, 15/02/2017, min:02:45:17).

Another expert also mentioned that there was more density in the peatland population. Population also covers the behaviour of people in the area. Since population density alone could not describe population behaviour, experts suggested having another node to cover this.

> “if the people stays there is civil servant they do not want to burn... using fire is dangerous, they do not want to do it”(E1, 15/02/2017, min:00:49:17).

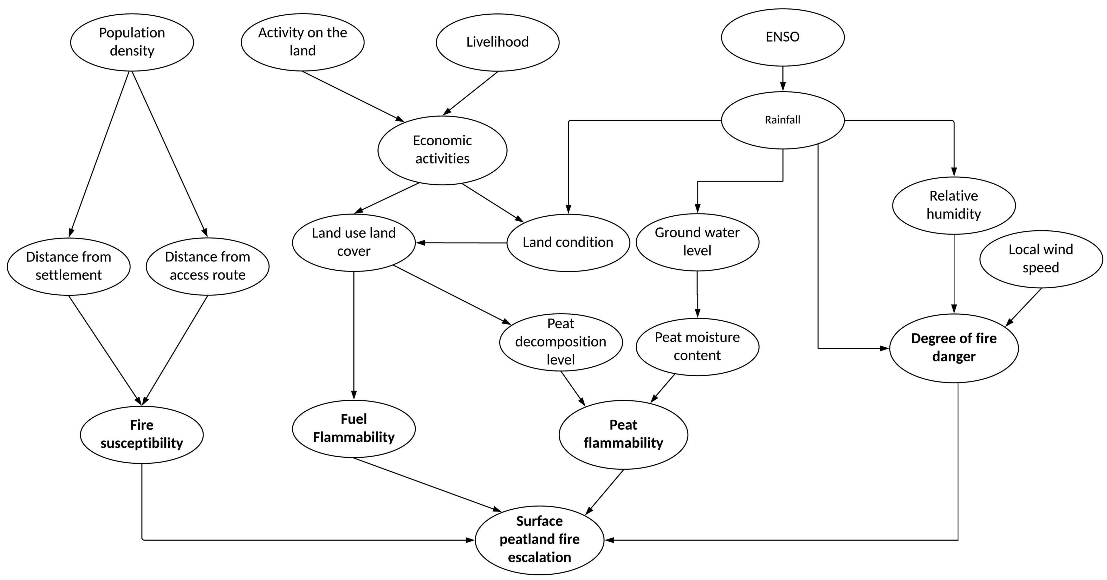


Figure 5.13: Version-3 of the structure of the causal model for hotspot escalation into peatland fires after the evaluation. This structure is a result of the evaluation, revised based on the input from the experts.

> "... here [an area] no people live, but more fire here. Many people here [an area], no fire... "(E6, 15/02/2017, min:00:50:30).

Therefore, a new node was introduced to cover the behaviour problem. This node was labeled *culture* and refers to the fact that even though an area has high density, if most of the people living there are Dayak people then they know how to manage the fire so it will not escalate (Vayda, 1999; Suyanto et al., 2009). Therefore, experts emphasised that having only population density could not represent the behaviour of people living in one area when they are using fire.

The new structure resulting from the second iteration is presented in Figure 5.13. The next step is determining the states for each node.

5.2.6 Determining the States of Each Node

Once the experts reached agreement on the nodes in the structure of the causal model, each node then should be discretised into states. In this research project, the states were established using recognised classification, available management thresholds, and expert knowledge. Expert knowledge was used for all nodes without clear classification or threshold.

The nodes that were discritised based on the recognised threshold were *rainfall*, *ground water level*, and *land use land cover*. The continuous values in the ground water level were classified into two states, deeper than 40cm from the surface and shallower than 40cm from the surface. This classification was proposed by Putra et al. (2011) and used by the Indonesian government as a valid threshold for ground water level. The continuous values for rainfall were also classified into two states, above 100mm and below 100mm. Based on the Aldrian and Dwi Susanto (2003), during the dry season the monthly amount of rainfall reaches below 100mm. It is also known that peatland fire incidents in Indonesia occurred almost every dry season (Harrison et al., 2009). The states for node *land use land cover* was also determined from the government classification (, SNI). However, for this causal model only the primary classes were used as the states.

Nodes without clear classification or valid threshold were discritised using expert knowledge. For example, there are many available options of livelihood that can be used as the states. However, the states should represent the livelihood around the study area. Based on the discussion with experts, around the study area there are few types of livelihood that influence the occurrence of peatland fire. This livelihood was then classified into five categories: agriculture, intentional fire escalation, timber harvesting, agroforestry, and non-timber forest production.

The detail of states for each node is presented in Table 5.3 and the description of each node together with the states is provided in the Appendix B.

5.2.7 D-separation Test for the Causal Model Structure

Causal networks also can be used to follow how a change of certainty in one variable may change the certainty for other variables (Nielsen and Jensen 2007). d-separation rules decided for each pair of variables, in the peatland fire causal model, whether they were independent given the evidence entered into the network. For example (see Figure 5.17), the nodes *rainfall* and *peat flammability* are D-separation by the node *ground water level*. This means that if the known ground water level is below the threshold, then having information that the amount of rainfall is less than 100 mm will not change the belief that the peat flammability will be high.

5.2.8 Refinement of the Structure - Third iteration

Another iteration of the refinement of the causal model structure was conducted in the evaluation workshop with the experts. The evaluation was conducted after the parameterisation process, in order to evaluate the performance of the causal model in capturing the characteristics of peatland fire escalation. During the evaluation process, the structure of the causal model was adjusted and refined based on the result of the survey for the parameterisation process. The changes to the causal model were based on the experts' evaluation and opinion. There are two significant changes to the structure of the causal model as result of the evaluation workshop. The first is removing some irrelevant nodes; the second is combining multiple nodes.

5.2.8.1 Removing Irrelevant Nodes and Links

During the second iteration of the structure development there was a node that rose an issue during the discussion. *Canal berm* is a node that for few experts is one of the important contributing factors. However, for some of the fire experts, the term *canal berm* is unfamiliar. One expert, E1, questioning the importance of canal berm as the contributing factors since the term *canal berm* is unfamiliar in the fire community in Indonesia. However one expert, E3, argued that based on the significant of the existence of *canal berm* as the contributing factors in peatland fire escalation, the node canal berm should be included in the structure of causal model.

> "... Why do you put canal berm there? This is for the first time, I heard this term [canal berm].... "(E1, 14/02/2017, min:02:35:40).

> "... canal berm is significant[in peatland fire] it has [drier] fuel [than the surrounding area] .... "(E3, 14/02/2017, min:02:36:37).

During the parameterisation process, when survey for eliciting the compatible parent configuration (CPC) form was sent to the expert, another issue arose. An expert did not fill the question related to canal berm. Based on our personal communication, this issues was clarified, the expert mentioned that the term *canal berm* is unfamiliar in peat science, therefore the expert can not answer the question.

> "The term of canal berm is not familiar in the peat science. I have been in this field for more than 20 years, and now I am part of the Peatland Restoration Agency, we never heard about that term... "(personal communication with E1, 22/04/2017).

In the evaluation workshop, this issue was raised and the fire experts once again were asked whether the node *canal berm* should be included in the causal model. Since this causal model might be implemented in the agencies and fire authorities in Indonesia, the term of *canal berm* could rise a bit confusion among the user of this causal model. Therefore, the node *canal berm* should be excluded from the causal model. Because this node is deleted from the structure of the causal model, the link from this node to its child node *fuel flammability* was also removed (see Figure 5.14).

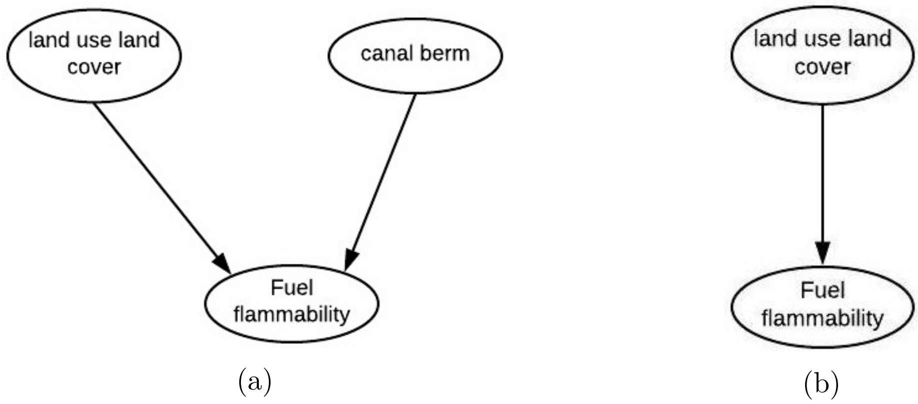


Figure 5.14: The new structure of fuel flammability. (a)fuel flammability before canal berm node was removed ; (b) fuel flammability after canal berm node was removed

5.2.8.2 Combination of Two Nodes

During the evaluation workshop, an expert mentioned brought up the possibility of missing information about planned burning conducted by the local people in Central Kalimantan. The culture of people around the peatland area should not be limited into causing or not causing the wildfire, but also on how the behaviour of people in conducting the planned burning. Local people has their own local wisdom in doing the prescribed or planned burning, while the foreigner most likely to be ignorance on this local wisdom. Therefore, node *culture* at this moment has not yet captured the planned burning.

Experts then threw different options on how to include the action of planned burning into the causal model, such as adding new states in the node culture, creating nodes that represent the ownership of the land.

> "... local people usually they done [planned burning] in one cday because the area is so small. But if it is a business, the area must be bigger.... recently because the change in behaviour, they let the fire grow up..."(E1, 3/05/2017, min:02:53:00).

> "... but should the culture has three categories, ... culture unlikely to use fire, most likely to use fire and more likely to use planned burning"(E4, 3/05/2017, min:02:54:00).

> "how about you have company land and private/public land?"(E4, 3/05/2017, min:02:54:00).

At first, nodes about land ownership was introduced to incorporate planned burning in the causal model. Three new nodes were introduced, *private land*, *leased land* and *government land*. The reasoning is private and public land were more likely to be look after, because it is owned by local people or local government. While leased land most likely has nothing to do with the culture and mostly under the ownership of big company, sometimes experienced illegal massive planned burning due to land clearing activities.

However, an argument happened in deciding the links of these nodes to the existing nodes. One expert was saying that the ownership of private land should influence the livelihood, another expert disagree on this scenario. Another argument also about the link of public land to the economic activity, one expert was questioning the reasoning of having direct link from public land to economic activity. This argument then solved when an expert suggested to combine the three nodes, and incorporated these into the node *culture*.

The node *culture* was renamed and redefined as *activity on the land*. Under the new definition of node *activity on the land*, the existence of node human action become *redundant*. These two nodes, both represent the actions or activities that taken by the people around the peatland area. Therefore, the node human action was removed, and a direct link was drew between node activity on land and economic activity (see Figure 5.15(b)).

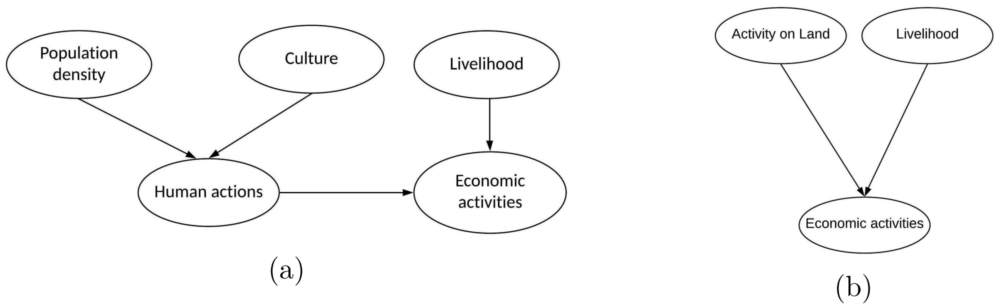


Figure 5.15: The new relationship on the *economic activity* node. (a)before deleting *culture* and *human action* node rainfall ; (b) after changing *culture* to *activity on land* node

The evaluation workshop has resulted a new structure of causal model as shown in Figure 5.13). A self-walkthrough model evaluation was conducted on this new structure. While doing this walkthrough evaluation, an interesting information was found about *ground water level* that make me have to refine the structure of the causal model. Further explanation on this structure refinement can be found in the next subsection.

5.2.8.3 Adding Relevant Nodes and Links Based on Literature Review

In the causal model presented in Figure 5.13, there is a node labeled as *rainfall*. This node is described as the summation of monthly rainfall expressed as daily rainfall. However, Putra et al. (2016) describes the condition of ground water level as influenced by the rainfall in the previous month. Their finding shows that in the year 2011-2012 the lowest precipitation occurred in August, and one month after, which is September, the ground water level dropped to the lowest point. This one-month lag time in the rainfall should be considered as a factor influencing the ground water level and not the current condition of the rainfall. Therefore the relationship of the node *rainfall* and node *ground water level* in the current causal model does not match with the rainfall condition referred to in Putra et al. (2016). The direct link from node *rainfall* to the *node ground water level* should be removed. Instead, a node representing a one-month time lag of rainfall should be created and a direct link from this node to node *ground water level* also needs to be drawn. Therefore, the structure of the causal model was modified by adding a new node, labelled as *summation of monthly rainfall-previous 30 days*, and created a direct link from this node to the *ground water level* node (see Figure 5.17).

Another change in the structure of the causal model is made in node *peat moisture content*. Previously, node *peat moisture content* only has one parent node which is the ground water level (see Figure 5.16(a)). This direct link was made based on a suggestion from the experts (described in Subsection 5.2.5) and supported by much research that has found a strong correlation: when the ground water level is low, the peat moisture content is also low (Usup et al., 2004; Wooster et al., 2012; Putra et al., 2016). However, it is also found that the temporal rainfall especially in a peatland area reduces the flammability of the area. If there is heavy rainfall on a particular day, the number of fire hotspots usually decrease on the next day (Putra et al., 2011). This temporal rainfall most likely creates a moist condition on the surface of the peatland. Therefore, it is unlikely that a fire hotspot escalates into wildfire under this moist condition. Based on this explanation, adding a direct link from the *rainfall* to *peat moisture content* was considered.

The final structure of the causal model is presented in Figure 5.17. Due to the changes in the structure of the causal model, another parameterisation process is needed for nodes that are affected by the changes. The process of extracting the new CPT for the affected nodes is presented in Chapter 6.

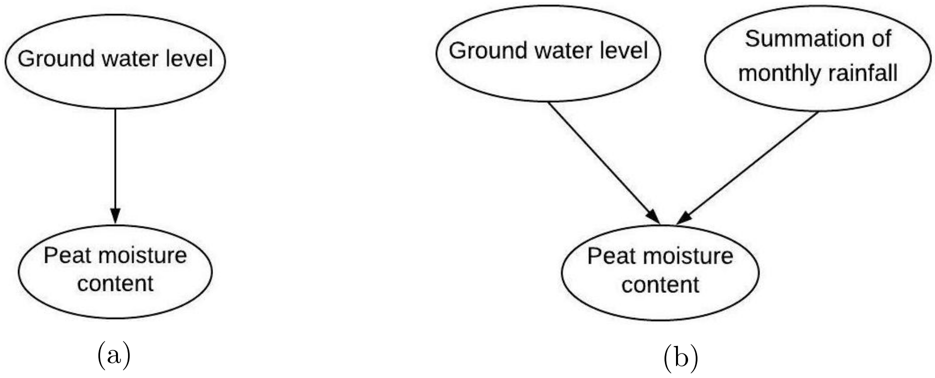


Figure 5.16: The new structure of the peat moisture content node. (a) before adding node rainfall; (b) after adding node rainfall

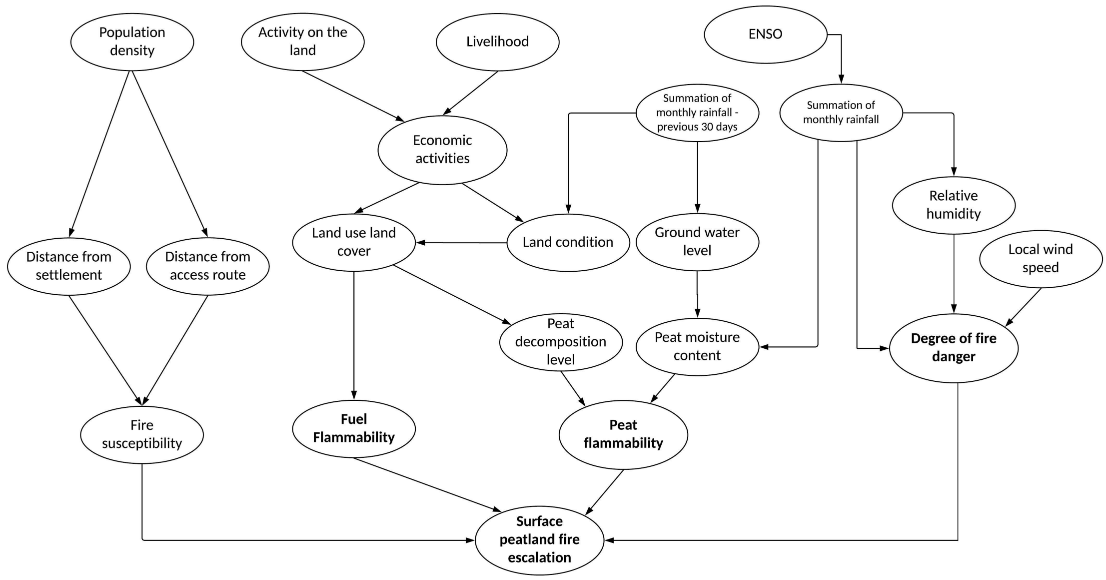


Figure 5.17: Version-4 of the structure of the causal model for hotspot escalation into peatland fires

5.3 Reflections on the Dynamics of Focus Group Discussion

In this research project, a focus group discussion was chosen as the method to elicit the expert knowledge. As explained in the literature review, Chapter 2, the group dynamic in a focus group discussion might influence the experts' contribution. There are two possible factors influencing the group dynamic discussed in this chapter: internal factors and external factors.

5.3.1 Internal Factors

The dynamic of a focus group discussion is dependent on experts involved in the session. The expert's background knowledge and capability to provide relevant information are two of the keys to obtaining a better result.

- Comprehensive nodes were gathered. Due to the diversity of expertise and knowledge from the experts, the content of knowledge bases is more complete. Most of the subdomains in the contributing factors of peatland fire escalation were represented at least by one expert.

- Less biases. Because the ideas from one expert can be clarified during discussion involving other experts.

- Lack of equal participation by all members (due to language barriers). In this research project, the group of experts involved in the focus discussion was made from three different nationalities. The discussion was conducted in English. Some difficulties in expressing thoughts were experienced by the experts due to English being a second language.

- Expert availability. To get all experts in the same room at the same time was also a challenge in this research project. Few of the participants were academic. The focus group discussion was conducted during the semester. One expert was absent for one day due to teaching responsibilities.

- Inconsistent input. During the discussion, not all the experts provided the input as expected. One expert tended to give or present his opinions on one topic when the discussion of the topic finished. This action sometimes delayed the progress of elicitation process because at some points the discussion had to come back to the previous topics.

- Difficulties in achieving consensus. For example in this research project, there are challenges in deciding the objective of the causal model. We spent one day just arguing and deciding what is the objective. Each expert has their own opinion about the importance of a problem.

5.3.2 External Factors

The external factors that might influence the dynamic in a focus group discussion sometimes cannot be seen and predicted before the session starts. These factors could influence the behaviour of the experts participating in the discussion.

- Having the list of causal variables presented in earlier in the discussion had made the knowledge elicitation process more focused.

- Time consuming and expensive. This is despite much research believing that focus group discussions are less time consuming compared to individual interviews (Krueger, 2014) and might lead to a lower cost. However, in this research project, due to the complexity of the problem and variety of expert expertise, the focus group discussion was conducted in few iterations in different days. The cost to invite experts from different regions and countries also was not cheap. However, this challenge could be accommodated by having *Skype meetings* with the experts.

- Jet lag. A few experts involved in the focus group discussion arrived from overseas one day before the focus group discussion. Adapting to the time difference between the host country and their original country was not an easy job for a few participants in the focus group discussion. In this research project, one expert could not focus during the discussion due to the jet lag experienced.

5.4 Chapter Summary

In this chapter, the second stage of the workflow for developing the causal model is presented. The knowledge from domain experts is used to build the structure of the causal model, starting with defining the objective of the causal model, identifying the causal variables, determining the relationship for each variable, and evaluating the structure of the causal model. All of these processes are conducted together with the domain experts through a focus group discussion. Since the causal model is represented by a graphical model using BN, the evaluation of the structure of the causal model follows some of the BN guideline.

This chapter also addresses Research Question 2.2 on how to incorporate expert knowledge to build a comprehensive causal model for surface peatland fire escalation. Experts with different expertise and experience were invited to join the focus group discussion. The structure of the surface peatland fire escalation causal model was constructed through focus group discussion and a few iterations of evaluation.

During the process of expert knowledge elicitation using focus group discussion, a few challenges arose and might have impacted the result of the process. However, despite all of these challenges, a comprehensive causal model for explaining the surface peatland fire escalation was established. The next stage was to parameterise the structure of the causal model.

Table 5.3: Nodes and states in peatland fire escalation causal model

Variables	**States**	**Source**
Population density	Low, Moderate, High	Statistic Indonesia
Distance of hotspot from the nearest settlement	Less than 1 km Between 1 km - 4 km More than 4 km	Expert knowledge
Distance of hotspot from the nearest access route	Less than 1 km Between 1 km - 4 km More than 4 km	Expert knowledge
Culture	Less likely to use fire More likely to use fire	Expert knowledge
Livelihood	Agriculture Intentional fire escalation Timber harvesting Agroforestry Non timber forest product	Expert knowledge
Human action	Low, Moderate, High	Expert knowledge
Economic activities	Low activities Medium activities High activities	Expert knowledge
Land condition	Degraded Non-degraded	Expert knowledge
Land use and land cover	Agriculture land Mixed forest/mixed shrub Open area Settlement Water body	the Indonesia National Standardization Agency (, SNI)
Canal berm	Yes, No	Expert knowledge
Rainfall	Below 100mm Above 100mm	Threshold
Relative humidity	Low, Medium, High	Expert knowledge
Local wind speed	Low, High	Expert knowledge
Ground water level	Deeper than 40 cm from the surface Shallower than 40 cm from the surface	Threshold (Putra et al., 2011)
Peat decomposition level	Sapric, Hemic, Fibric	Guideline
Peat moisture content	Low, Medium, High	Expert knowledge
Fire susceptibility	Low, Medium, High	Expert knowledge
Fuel flammability	Low, Medium, High	Expert knowledge
Peat flammability	Low, Medium, High	Expert knowledge
Degree of fire danger	Low, Medium, High	Expert knowledge
Surface peatland fire escalation	Yes, No	Expert knowledge

Table 5.4: The final version of nodes and states in peatland fire escalation causal model

Variables	States	Source
Population density	Low, Moderate, High	Statistic Indonesia
Distance of hotspot from the nearest settlement	Less than 1 km Between 1 km - 4 km More than 4 km	Expert knowledge
Distance of hotspot from the nearest access route	Less than 1 km Between 1 km - 4 km More than 4 km	Expert knowledge
Activity on land	Less likely to use fire More likely to use fire	Expert knowledge
Livelihood	Agriculture Intentional fire escalation Timber harvesting Agroforestry Non-timber forest product	Expert knowledge
Economic activities	Low activities Medium activities High activities	Expert knowledge
Land condition	Degraded Non-degraded	Expert knowledge
Land use and land cover	Agriculture land Mixed forest/mixed shrub Open area Settlement Water body	the Indonesia National Standardization Agency (, SNI)
Summation of 30-days Rainfall	Below 100mm Above 100mm	Threshold
Summation of 30-days Rainfall in previous month	Below 100mm Above 100mm	Threshold
Relative humidity	Low, Medium, High	Expert knowledge
Local wind speed	Low, High	Expert knowledge
Ground water level	Deeper than 40 cm from the surface Shallower than 40 cm from the surface	Threshold (Putra et al., 2011)
Peat decomposition level	Sapric, Hemic, Fibric	Guideline
Peat moisture content	Low, Medium, High	Expert knowledge
Fire susceptibility	Low, Medium, High	Expert knowledge
Fuel flammability	Low, Medium, High	Expert knowledge
Peat flammability	Low, Medium, High	Expert knowledge
Degree of fire danger	Low, Medium, High	Expert knowledge
Surface peatland fire escalation	Yes, No	Expert knowledge

Chapter 6

Parameterisation of the Peatland Fire Escalation Causal Model

Introduction

This chapter introduces the third stage of the workflow of causal model development. This third stage aims to elicit the probability distribution of each node in the causal model and generate the conditional probability tables (CPTs). The probability distribution is elicited from the expert knowledge and an online survey is set up to obtain the experts' answers. In this research project, the method proposed by Das (2004) is adopted to reduce the number of questions about the probability distribution. The Analytic Hierarchy Process (AHP) (Saaty, 2008) is also implemented to perform the weight comparison between experts' answers and also to accommodate the different knowledge backgrounds and expertise of the experts.

This chapter discusses the implementation of this third stage in the development of the peatland fire escalation causal model. A section about what has been learned from the implementation of this process is also presented, followed by a summary of the chapter.

There are two ways to elicit the probabilities using expert knowledge, direct elicitation, and indirect elicitation. Direct elicitation is mostly used for discrete variables, while indirect elicitation is used for continuous variables. In the direct elicitation approach, experts should express their degree of belief as a number. It can be a probability, a frequency, or a ratio (Renooij, 2001). In the indirect elicitation approach, experts are asked for a decision from which their belief is inferred. In this research project,the focus is on the direct elicitation approach using a probability-scale method.

A probability-scale method is a well-known direct method, where experts are asked to indicate their belief on a scale. The scale could be either 0 - 1 or 0 - 100%. In addition to the probability scale, (van der Gaag et al., 1999) described a probability scale with numerical and verbal anchors (see Figure 6.1). The numerical scale range from 0% to 100% and the verbal cues in the scale included *(almost) certain*, *probable*, *expected*, *fifty-fifty*, *uncertain*, *improbable*, and *(almost) impossible*.

For each probability that is to be assessed the expert is asked to choose the "correct" number or verbal description. A figure containing a description of the required probability and scale is presented to the experts. The use of number probability of verbal description depends on how familiar the experts are with the probability to be assessed (Renooij, 2001). Even though the assessments using a probability scale tend to be inaccurate and prone to bias (source), this method is easy to understand and use. It provides a fast way of elicitation especially for elicitation of large number probabilities (Renooij, 2001).

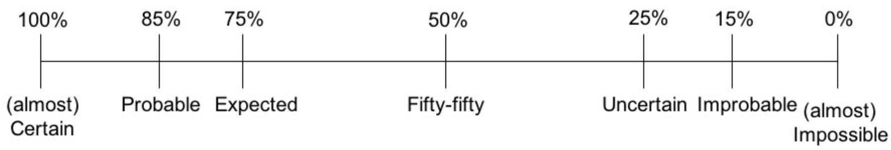


Figure 6.1: Probability scale, adapted from van der Gaag et al. (1999)

6.1 Generating Questions for the Online Survey

This section discusses how to create questions for eliciting the probabilities from experts. The probabilities elicited from the experts are used to populate the CPTs. Experts are asked to assess a scenario and express the degree of their belief. The CPTs for the hotspot escalation into peatland fire causal model is generated using the fire experts' knowledge. The same fire experts involved in the workshop of structure development are invited in this process (see Table 3.1).

A set of questions representing the scenario is generated based on parent nodes and the states. The number of questions asked depends on the complexity of the causal model. The more parent nodes and states of a causal model, the more questions are generated and asked. The questions for eliciting the probability distribution can be categorised into three different categories: questions for nodes without parents, questions for a node with single parent, and questions for a node with multiple parents.

The online survey is set up using the Survey Monkey platform. An email invitation is sent to experts to let them assess the survey. This survey remains open for two weeks and the experts can enter answers in their own time. Experts are also able to change their answers after submitting the survey, as long it is within the period of two weeks.

6.1.1 Questions for Nodes Without Parents

In a Bayesian Network, a node without parents is categorised by the marginal probability distribution of its states. The number of questions to elicit the marginal probability depends on the number of states present in the node. The questions focus only on the probability of each state. The template that can be used to elicit the probability is presented below:

> What is the likelihood of the following scenario:
>
> **State**

State in the question should represent the state of the node, it can be described and modified so it is more coherent and understandable by the experts.

In the structure of hotspot escalation into peatland fire causal model, there are a few nodes with no parent. For some of the nodes, the marginal probabilities can be generated based on the historical data such as *population density* and *ENSO*. However, a few nodes do not have historical data such as *culture* and *livelihood*. The probabilities for these nodes are then obtained from the experts. A set of questions is generated to elicit the probabilities.

Consider the node *livelihood* as an example of a question created for a node without a parent. This node has five states. Five questions are created in association with each state in the nodes. Experts are asked to provide their belief in the probability for each state in the node *livelihood*. Figure 6.2 shows one question created for the state *agriculture* in node *livelihood*. In this question, experts are asked to choose one probability that is displayed as multiple choice. The options in the multiple choice are generated based on the probability-scale method, as shown in Figure 6.1. There is an additional option of *other*. The experts could fill this option if they have another probability that has not mentioned in the choices.

6.1.2 Question for Node With Single Parent

The conditional probability distribution is obtained from a node with single/multiple parents. For a node with a single parent, the question to elicit the conditional probabilities is similar to the question asked for a node without parents. The only difference is that the probability being elicited is associated with the states of the parent node. The question is generated for each state in the child node followed with the condition of the states in the parent nodes. Below is the template for a question to be asked for a node with a single parent

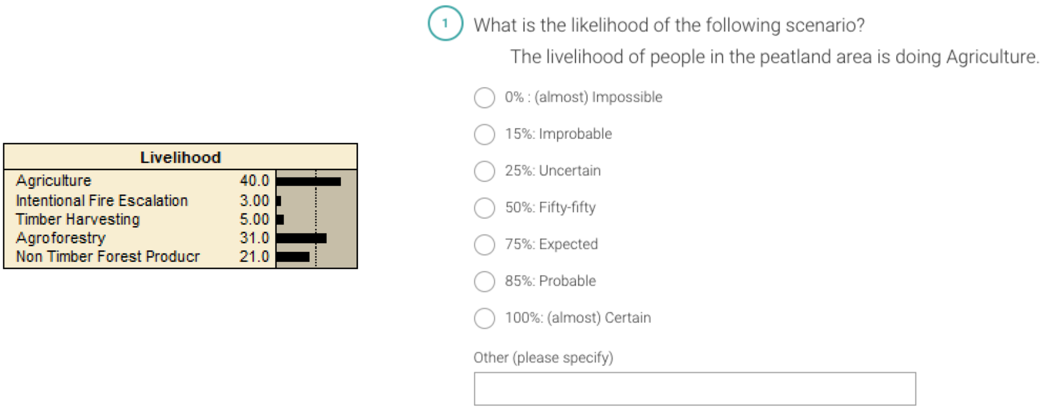


Figure 6.2: A sample question for a node without a parent

What is the likelihood of the following scenario:

ChildState

If we know that:

ParentState

In the causal model for predicting the escalation of hotspots into peatland fires, there are few nodes that have only single parent. For example node *distance of hotspot from the nearest settlement*. This node has only one parent node, *population density* with three states (low, moderate, and high). The node *distance of hotspot from the nearest settlement* also has three states (less than 1km, between 1km - 4km, and more that 4 km). The questions created to generate the CPT for this child node are based on each state of the child node in association with each state of the parent node. In order to generate the CPT of this child node, nine questions are created. This means there is a question for every state in node *distance of hotspot from the nearest settlement* in association with every state in node *population density*. Figure 6.3 shows one sample question created for the state *less than 1 km* of the child node *distance of hotspot from the nearest settlement*, in association with state *low* in the parent node *population density*.

Similar with the question for nodes without parents, experts can choose one of the probability options or put their own answer in *other* option.

6.1.3 Question for Node With Multiple Parents

In the causal model for predicting hotspot escalation into peatland fires, there are a few nodes with multiple parents. These nodes also have a varied number states, from two to five states. Due to a varied number of parent nodes and states for each child node, the number of parameters to assess becomes large. The more parameters

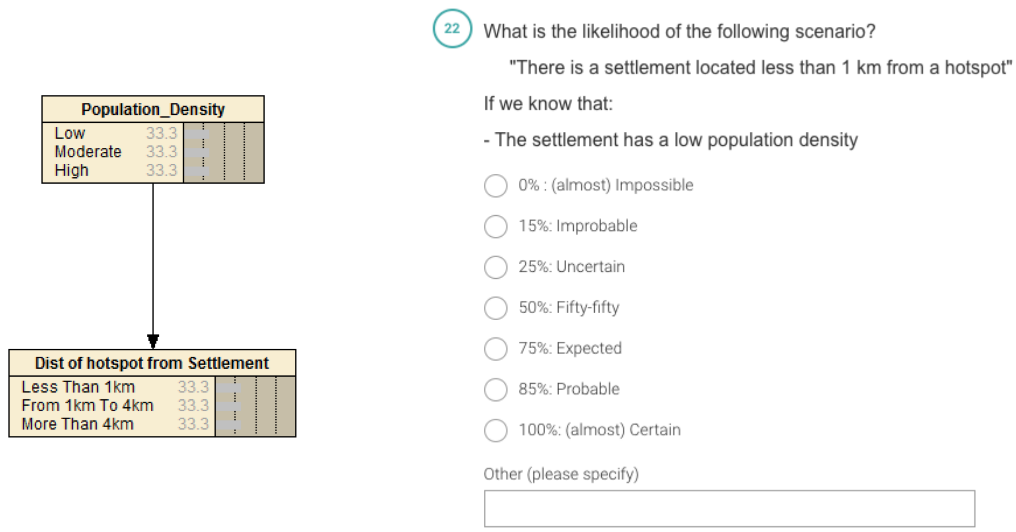


Figure 6.3: A sample question for a node with a single parent

to be assessed, the more questions needed to populate the CPTs. As mentioned in the literature review Section 2.4.1.1, answering a large number of questions, which is sometimes a bit repetitive, could stress out the experts and influence their answer. Therefore, a weighted sum algorithm proposed by Das (2004) is used to populate the CPTs. Using this algorithm the number of questions asked of experts becomes lesser and could reduce the burden on experts in answering the question.

In order to describe the process of populating the CPTs using weighted sum algorithm from Das (2004), consider child node *surface fire escalation (SFE)* from the causal model (see Figure 6.4). This child node has four parent nodes (*fire susceptibility*(*FS*), *fuel flammability* (*FF*)), *peat flamability (PF)* , and *the degree of fire danger (FDI)*. The child node *SFE* has two states, *No* and *Yes*. The parent nodes have three states, *low*, *medium/moderate*, and *high*. The CPT for *SFE* comprises 4^3 or 64 probability distributions. This means for each state in the *SFE* node, we need to create 64 questions. Since there are two states in this node (*No* and *Yes*), the number of questions that need to be created to populate the CPT is 128 questions. This number of questions could be reduced, which requires less effort from the experts to answer the questions.

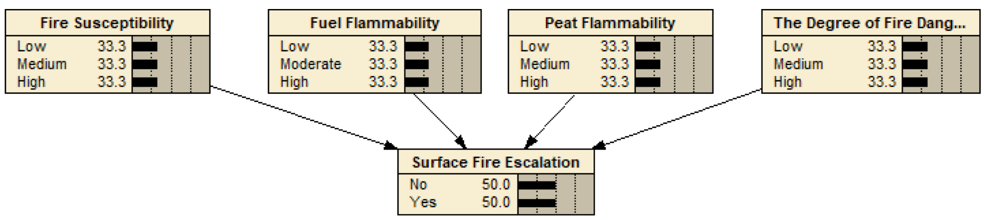
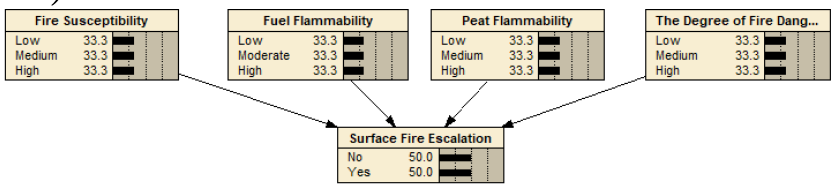


Figure 6.4: A network to assess the escalation of hotspots into peatland fires.

The process to create the question starts by asking the experts to choose the most compatible states of a parent node to the states in other parent nodes. The example of question for choosing the CPC for *SFE* nodes is shown in Table 6.1. In this question, experts were asked to choose the state in one parent node that is most likely compatible with states in the other parent nodes.

Table 6.1: CPC question for sub model fire susceptibility

For each state of each parent, which state of each other parent is ***most likely (compatible)****?*



FS	**FF**	**PF**	**FDI**	**FF**	**FS**	**PF**	**FDI**
low				low			
medium				medium			
high				high			
PF	**FS**	**FF**	**FDI**	**FDI**	**FS**	**FF**	**PF**
low				low			
medium				medium			
high				high			

The experts' responses are then gathered. All of the experts agreed that the state *low* in node *FS* was equivalent to state *low* in nodes *FF*, *PF*, and *FDI*. The state *medium* in node *FS* was equivalent to state *medium* in nodes *FF*, *PF*, and *FDI*. The state *high* in node *FS* was equivalent to state *high* in nodes *FF*, *PF*, and *FDI*. Based on these responses, the CPCs for *SFE* are created.

- CPC(FS = low) ≡ {FS = low, FF = low, PF = low, FDI = low}
- CPC(FS = medium) ≡ {FS = medium, FF = moderate, PF = medium, FDI = medium}
- CPC(FS = high) ≡ {FS = high, FF = high, PF = high, FDI = high}

Once the relevant CPCs have been elicited, then the questions for eliciting the conditional probability of the child node states can be generated using all of the possible CPCs. One question is created for each CPC in association with each state in the child node, *fire susceptibility*. Therefore, instead of answering 128 questions in order to populate the CPT for node *fire susceptibility*, the experts only need to answer six questions.

Figure 6.5 shows one of the questions generated based on the CPC(FS = low) for state = *Yes* of child node *SFE*. The probability-scale method as shown in Figure 6.1 is used to indicate the probability that could be chosen by the experts. The complete details about CPC questions and the online survey questions to elicit the conditional probability can be found in the Appendix C.

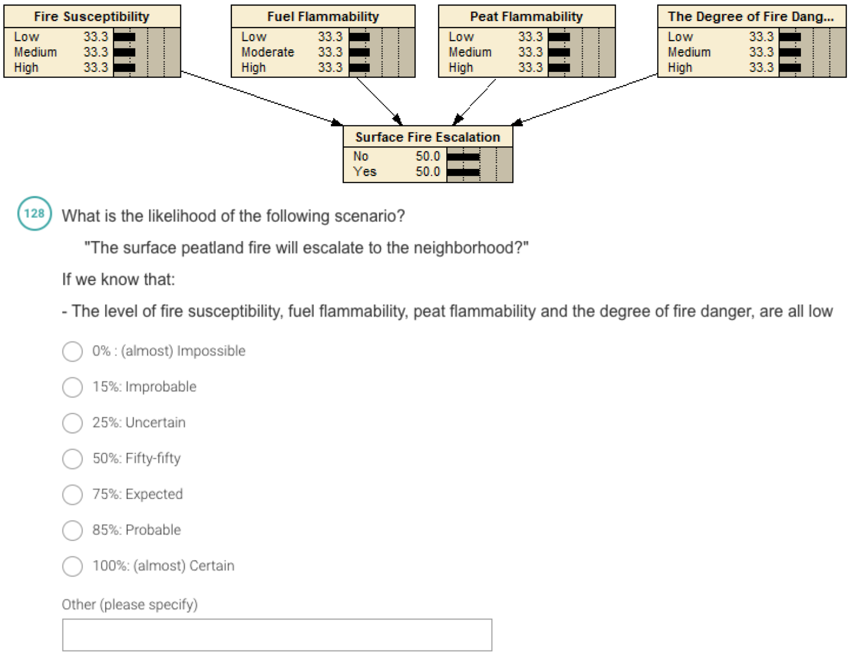


Figure 6.5: Sample question for a child node with multiple parents

6.2 Calculating Conditional Probability Tables (CPTs)

Once all the probabilities are gathered, the next information needed to populate the CPTs contains the relative weight of each parent of a child node. Another set of questions is set up to obtain the relative weight of each parent from the experts. In the questions, experts are asked to do a pairwise comparison of each parent and to choose which one is more influential (or equally influential) to the child node.

Let consider the same sub-model in Subsection 6.1.3 as the example for this CPT calculation. Experts are asked to choose which parent node more greatly influences another parent. In this example, experts have to choose the one that has more influence in hotspot escalation into peatland fire, between *fuel flammability* and *peat flammability*; *fuel flammability* and *degree of fire danger*; *fuel flammability* and *fire susceptibility*; *peat flammability* and *degree of fire danger*; *peat flammability* and *fire susceptibility*; *degree of fire danger* and *fire susceptibility*. Choosing a number between 1 - 9 is required to show how much more influential a parent node is to other parent nodes. Table 6.2) shows a respond from one expert on the pairwise comparison of each parents for *surface fire escalation node* . The numbers were derived from the AHP process documented in (Saaty, 1990). If the experts think that both parents have the same influence, they can enter the same numbers. The complete set of questions to obtain the relative weights for each CPC is provided in Appendix C.2.

Table 6.2: Relative weight question

Which parent variable (fire susceptibility, fuel/degree of flammability, peat flammability, or fire danger index) has the largest influence on surface fire escalation ? *How much more influential is this parent node against one of the other parents? (1 - 9 times)*			
Fuel flammability	1	**Peat flammability**	6
Fuel flammability	5	**Degree of fire danger**	1
Fuel flammability	1	**Fire susceptibility**	6
Peat flammability	6	**Degree of fire danger**	1
Peat flammability	6	**Fire susceptibility**	1
Degree of fire danger	6	**Fire susceptibility**	1

The result of the relative weights from each expert transfers into a matrix. A principal of eigenvector (Saaty, 1990) is used to process the matrix mathematically to get the relative weight. Table 6.3 shows the matrix and the final relative weight.

Table 6.3: Calculation of relative weight using eigenvectors

	FS	**FF**	**PF**	**FDI**	**Weight**
FS	0.075	0.454	0.111	0.013	0.163
FF	0.012	0.075	0.111	0.410958904	0.152
PF	0.455	0.454	0.667	0.493150685	0.517
FDI	0.455	0.015	0.111	0.082	0.166

Table 6.4: Distribution over node *surface fire escalation* for compatible parental configurations $Comp(FS = low)$

Probability distribution over *SFE*	s = low	s = medium	s = high
$p(SFE = no \mid Comp(FS = s))$	0.75	0.5	0.15
$p(SFE = yes \mid Comp(FS = s))$	0.15	0.25	0.85

Once all the information was obtained such as the relative weight (see Table 6.3) and the probability distribution over a child node (see Table 6.4), the CPT can be generated. However, the sum of the probability distribution over SFE for each e was not equal to one. Therefore, this probability distribution needs to be normalised before inserting the calculation of CPTs. After the two distributions in Table 6.4 are normalised, all the 4^3 distributions required to populate the CPT could be calculated.

Let consider a simple case where the distribution over *SFE* is calculated when the *FS*, *FF*, *PF*, and *FDI* are all in the state $s = low$. Based on the Equation 2.9, the probability distribution over child node *SFE* with the state $s = low$, i.e:

$$\{p(SFE = e|FS = low, DA =< 1km), e = no, yes\} \tag{6.1}$$

According to the equation 6.1, the following computations need to be carried out:

$$\begin{aligned}
&p(SFE = e|FS = low, FF = low, PF = low, FDI = low) = \\
&w_1 p(SFE = e|\{Comp(FS = low)\}) + w_2 p(SFE = e|\{Comp(FF = low)\}) \\
&+w_3 p(SFE = e|\{Comp(PF = low)\})+w_4 p(SFE = e|\{Comp(FDI = low)\}), \\
&e = no, yes
\end{aligned} \tag{6.2}$$

The experts assigned relative weights $w_1 = 0.163$, $w_2 = 0.152$, $w_3 = 0.517$ and $w_4 = 0.166$ for the nodes *FS*, *FF*, *PF*, and *FDI*, respectively. Using the relative weight assigned by the experts, the computation in Equation 6.2 gives:

$$\begin{aligned}
&p(SFE = e|FS = low, FF = low, PF = low, FDI = low) = \\
&0.163p(SFE = e|\{Comp(FS = low)\})+0.152p(SFE = e|\{Comp(FF = low)\}) \\
&+0.517p(SFE = e|\{Comp(PF = low)\})+0.166p(SFE = e|\{Comp(FDI = low)\}), \\
&e = no, yes
\end{aligned} \tag{6.3}$$

A weighted sum algorithm described in Chapter 2 subection 2.4.1 was implemented to estimate the rest of the probabilities in the CPT.

6.3 Combining the Conditional Probability Tables (CPTs) from Multiple Experts

Before analysing the result of the causal model, the CPTs generated by each expert must first be collated into a single CPT. In this research project, seven experts were invited to fill the online survey. Six experts had been able to complete the survey and were labeled as E1 - E6; one expert did not finish the survey. Therefore, in the process of collating the experts' answers, only the complete answers from six experts have been taken into account.

In this research project, the CPTs from multiple experts need to be combined into a single CPT. In order to generate a single CPT, a weighted-mean calculation is used. A specific weight was assigned to each expert based on their background knowledge and experience. The AHP method explained in Chapter 2 subection 2.5.2 was implemented to calculate the weight of each expert.

6.3.1 AHP for calculating the Weight of each Expert

As mentioned in Subsection 2.5.2, this research implemented the AHP method to calculate the weight of each expert. The weight is generated for four sub-models: climatic conditions, human involvement, forest ecology, and peatland ecology.

Establish the hierarchy of the structure.

There are three levels of hierarchy created to determine the influence of each expert on the CPT calculation for the surface peatland fire escalation causal model (Figure 6.6. The first or top level refers to the overall goal of finding the weight of each expert. The intermediate level comprises the three criteria which contribute to the goal. In this thesis, three criteria to weight the experts were determined: work experience, local knowledge, and expertise. The third level contains the experts which are given weights based on the criteria in the intermediate level.

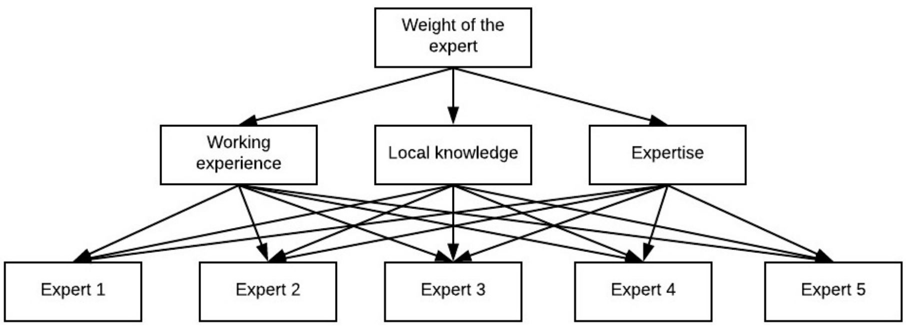


Figure 6.6: The hierarchy of structure in calculating the experts' weights

Elicit the pairwise comparison judgments.

In this step, all the criteria in the intermediate level are arranged into a matrix and elicit judgment of the relative importance of the criteria with respect to the overall goal. In this thesis, three pairs of criteria were established and compared. The first pair is work experience and local knowledge; the second pair is work experience and expertise; and the third pair is local knowledge and expertise. Each criterion in the pair then were given a weight based on the relative importance scale. For

Table 6.5: Eigenvector matrix for the criteria in the intermediate level

	Work experience	Local knowledge	Expertise	Priority
Work experience	1	1/3	1/5	0.105
Local Knowledge	3	1	1/3	0.258
Expertise	5	3	1	0.637
Consistency Ratio (CR) = 0.04				

example, the pair of works experience and local knowledge. Local knowledge were given a weight of 3, which means the local knowledge is judged as moderately more important than working experience. The detailed explanation of comparison of each pair and the weight that generates the criteria can be found in the Appendix D.

Based on the calculation of the eigenvector of the matrix (see Table 6.5), the criteria expertise has the highest priority with 63.7% influence, followed by local knowledge and working experience with 25.8% and 10.5% influence, respectively.

After the pairwise comparison matrix for the intermediate level is done, next is to generate the pairwise comparison of elements in the lowest level. The elements to be compared pairwise are the experts with respect to how much better one is than the other. This comparison is in association of how satisfying each criterion in the intermediate level. Thus there will be six matrices of judgments which are work experience, local knowledge, expertise on climatic condition, expertise on human involvement study, expertise on fuel and forest ecology, and expertise on peat ecology.

Since there are six experts that contributed to the online survey, the pairwise comparison matrix was set as 6 x 6. The judgment of each matrix relied on the assumption that was made based on the summary of experts' work experience with forest fire, experts' direct experience with forest fire in Indonesia, and experts' expertise in different domains.

For the pairwise comparison of experts' working experience, E1, E2, and E5 shared the highest priority, which is 29.4%. This judgment was based on the fact that these experts have work experience spanning more than 20 years in the forest fire field. Expert E3, E5, and E6 have a lower priority due to having less than 10 years work experience.

The results of pairwise comparison of experts' local knowledge follows. E1 has the highest priority due to being a local fire expert originally from Central Kalimantan, the study area, and having a lot of experience working with the local communities in this area. E5 and E6 are also originally from Central Kalimantan. However, both of the experts have less experience working in the fire communities compared to the other experts. E4 was given the lowest weight in this criteria because of limited experience with the fire conditions in Indonesia.

In the pairwise comparison of experts' expertise, there are four different weights that were assigned to each expert. These weights were categorised based on experts' expertise in climatic condition, peat ecology, fuel flammability, and human access.

A further description of the summary and judgment of each expert's comparison can be found in the Appendix D.

Establish the composite or global priorities of the lowest level with respect to the goal

The goal of this AHP calculation is to obtain the weight of each expert of different expertise. As mentioned above, there are four categories of expertise: climatic, human access, peat flammability, and fuel flammability. These categories are in association with the nodes and their relationship in the causal model.

The calculation of the weight is:

- The priority with respect to the work experience of each expert, multiplied by the work experience's priority and
- The priority with respect to the local knowledge of each expert, multiplied by the local knowledge's priority and
- The priority with respect to the expertise of each expert for each category, multiplied by the expertise's priority
- Sum up the total of the calculations above.

The result from the final weight calculation reveals that:

- For the climatic condition category, E4 has the highest weight followed by E6. These two experts were experts in the area related to climatic conditions during the forest fire. The judgment of the expertise was also made based on the fact that these two experts have publications about forest fires from the point of view of climatic conditions.
- For the human activity category, E5 has the highest weight with 50.4% influence and the gap is quite significant compared to the weight from other experts. This judgment was made based on the fact that this expert has worked with local people and fire communities in Central Kalimantan for a long time.

- For the fuel flammability category, E1 has the highest weight with 43.6% influence, followed by E5 with 24%. The judgment was made based on the fact that E1 has a formal education about forest fire ecology and has published many papers related to the forest fire from the perspective of ecology.

- For the peat flammability category, expert E5 has the highest weight with 33.3% influence, followed by expert E1 and expert E2 with 29.4% and 24.1% respectively. Expert E5 was assigned the highest weight due to experience working in the peatland area in Central Kalimantan. As local researcher, this expert has more local wisdom related to peatland fire compared to other experts.

The detailed calculation of the final weight can be found in the Appendix D.

6.3.2 Weighted-Mean for Combining the Conditional Probability Tables (CPTs)

Once the weight is estimated and assigned to each expert, the multiple CPTs for each node can be aggregated into a single CPT. In this research project, the weighted-mean method was used to combine the CPTs. In Subsection 6.3.1, four categories were set up to calculate the final weight for each expert. These categories are associated with the child nodes in the causal model (see Table 6.6).

Table 6.6: Classification of the nodes in the causal model with the categories of expert weight

Category	**Nodes**
Climatic condition	*Summation of 30-days rainfall, Relative humidity, Degree of fire danger*
Human activity	*Distance of a hotspot from the nearest settlement, Distance of a hotspot from the nearest access route, Economic activities, Fire susceptibility*
Fuel	*Land condition, Land use and land cover, Fuel flammability*
Peat	*Ground water level, Peat decomposition level, Peat moisture content, Peat flammability*

Consider the network in Figure 6.3. There are six different CPTs generated for the *dist of hotspot from settlement* node as a result of the responses from six different experts.These CPTs need to be aggregated into a single CPT before being included in the causal model. This node is under the category of human activities. Therefore the final weight for human activity was used to combine the CPTs.

6.4 Revision of CPTs for the Final Structure of the Causal Model

A description In Chapter 5, Subsection 5.2.8 shows a modification in the causal model that was made after this parameterisation process. Due to this modification, the CPTs of some nodes were affected. A new CPT for these node should be calculated. This section provides an explanation of how to generate/modify the CPTs for a node that is affected by the changes.

6.4.1 Generating New CPT for node *Peat Moisture Content (PMC)*

As shown in Chapter 5, Figure 5.16(a), node *peat moisture content* only has a single parent, while in the new structure as shown in Figure 5.16(b) node *PMC* now has two parents, *ground water level (GWL)* and *rainfall*. In the initial structure, node PMC only needs two probability distributions for each state (see Table 6.7). However, in the new structure the number of parents increases. The CPT for node *PMC* will comprise of four probability distributions for each state. Therefore, a new CPT for node PMC should be populated. Instead of sending a new survey to gather a new probability distribution, the probability distribution from the early survey is used to populate the CPT using the Das (2004) method .

Table 6.7: CPT of node *PMC* before the changes on the structure of causal model

Ground water level	low	moderate	high
Deeper than 40 cm from the surface	0.8	0.2	0
Shallower than 40 cm from the surface	0	0.46875	0.53125

The first step in populating the new CPT is creating the CPCs for node *PMC*. Based on the findings in the literature review and also information extracted from the focus group discussion, the state *below* in node *GWL* is considered compatible with the state below in node *rainfall*. The state *above* in node *GWL* is also most likely to coexist with the state *above* in node *rainfall*. Therefore, the CPCs for this node are presented below:

$$\{Comp(GWL = s)\} \equiv \{GWL = s, Rainfall = s\}, \text{for} s = below, above. \quad (6.4)$$

Based on Equation 6.4, the probability distribution over the child node PMC will have:

$$\{p(PMC = e|GWL = s, Rainfall = s), \text{for} s = below, aboveande = low, medium, high\} \quad (6.5)$$

Table 6.8: Distribution over node PMC for compatible parental configurations $\{Comp(GWL = s)\}$

Probability distribution over PMC	s = below	s = high
$p(PMC = low\|\{Comp(GWL = s)\})$	0.8	0
$p(PMC = medium\|\{Comp(GWL = s)\})$	0.2	0.46875
$p(PMC = high\|\{Comp(GWL = s)\})$	0	0.53125

The probability distribution in Table 6.8 was adjusted from the CPT in Table 6.7. The probability distribution over PMC for $p(PMC = e|\{Comp(GWL = below)\}), e = low, moderate, high$ was taken from the probability distribution of *gwl* = *below*. While the probability distribution over PMC for $p(PMC = e|\{Comp(GWL = above)\}), e = low, moderate, high$ was taken from the probability distribution of *gwl* = *above*.

The relative weights needed to populate the CPT using the weighted sum algorithm was set up as $w_1 = 0.75$ and $w_2 = 0.25$, for the parent nodes *GWL* and *rainfall* respectively.

6.5 Reflection on Eliciting Expert's Knowledge using Online Survey

Based on the responses from the experts through the online survey, a few issues were discovered. Most of the issues related to the questions that had been asked to the experts. These issues have affected the CPTs that were populated using the experts' responses. For some of the nodes, the probability distribution of each state became uniform and less distinctive.

In this section, the reflection on the issues when conducting the online survey for generating the CPTs is described. Some issues have been identified and explained in the list below:

1. Experts' biases when answering the questions

Experts invited to answer the online survey are the same experts involved in building the structure of the causal model. During the process of creating the structure of the causal model, the focus of discussion was to discover the factors influencing the escalation of hotspots into peatland fires. This focus seems to influence the experts' perspectives when answering the online survey. Experts were confident in answering the question related to the hotspot escalation, but indecisive in questions about non-escalation hotspots.

2. Language barrier, including negative sentences

   The questions in the survey were written in English. Half of the experts are non-english speakers. It is found that some experts have misunderstood the meaning of questions, especially for negative sentences.

3. Misunderstanding options in multiple choice questions

   The questions in the online survey were created as multiple choice questions. The multiple choice options were generated using the probability scale proposed by (van der Gaag et al., 1999). Experts mentioned that it is difficult for them to differentiate the meaning of each option in a multiple choice question. One expert specifically mentioned that options *uncertain* and *fifty-fifty* most likely have the same meaning. Even though the numerical scale of the probability was provided, the experts still had difficulties in choosing the most appropriate option.

4. Experts being indecisive

   A few experts were being indecisive in providing the probability for some of the questions. One expert chose the option 50% or fifty-fifty for most of the questions in the survey. This might indicate that the expert does not know the answer or does not have enough domain knowledge to elaborate on the question. The survey result also revealed that most of the experts were hesitant to provide a certain option, especially for the questions that were out of their expertise. For example, one expert with climatic condition expertise, was decisive in the question related to the climatic conditions. But, this expert was indecisive in the questions related to other areas, such as peatland ecology.

5. Design of online survey layout

   As explained in Section 6.2, the probability distribution given by the experts did not sum to one. The layout/format of the online survey, does not provide any information about the sum of the probability distribution over one node. Therefore, experts had no idea whether their answers fit the requirement of summing to equal one.

6.6 Chapter Summary

In this chapter, the third stage of the workflow for developing the causal model is presented. The variables/nodes and their relationships developed in the second stage were quantified. Experts' opinions were elicited in order to gather the probability distribution of each node in the causal model. The causal model in this thesis could be categorised as a complex model. The number of nodes was more than 20 and the relationship of each node was complicated. Therefore, the questions to be asked to populate the CPTs for the causal model are numerous. In order to ease an expert's burden in answering numerous questions, the Das (2004) method was implemented to populate the CPTs. By implementing this method, fewer questions were needed to populate the CPTs. The questions then were put in an online survey platform and sent to the experts.

During the process of the eliciting the experts' opinion on the probability distribution of each node, a few issues were identified and discussed. The issues are related to the design of the survey, that might experts' answers.

Chapter 7

Test Data Preparation, Test Result, and Analysis

The structure of the causal model has been developed in Chapter 5. The parameterisation process has also been completed in Chapter 6. The chapter covers the next stage in the workflow, to evaluate the performance of the causal model in capturing the complexity of peatland fires in Indonesia (see Figure 3.2). This chapter also aims to address RQ. 3 by evaluating the causal model when there is no gold standard model that can be used as comparison.

This chapter starts with the explanations of the study area and data preparation. The performance of the causal model is evaluated using a small data set of historical hotspot escalations that have been officially verified (ground checked) and a set of hotspot data inferred to be non-escalations based on the locations of the ground checked escalation data. The explanation of these test cases is provided in Subsection 7.2.2.

This escalation and non-escalation test data set is also used on an implementation of the published guidelines of determining hotspot escalation using satellite data from the Indonesian National Institute of Aeronautics and Space (LAPAN [1]) and the results are compared with the those of the causal model. The results of the evaluation are presented in Section 7.3. This section also covers the analysis on the most influences nodes in the structure of the causal model and examinations of the reasoning process of the causal model using four different hotspot scenarios as illustrations. The chapter concludes with a summary of the chapter.

[1]Lembaga Penerbangan dan Antariksa National

7.1 Study Area

The study area in this research is part of the ex-Mega Rice Project (Ex-MRP) area in the Kapuas district in Central Kalimantan which was also part of the Kalimantan Forest and Climate Partnership (KFCP) project area. This area covers approximately 120000 ha of tropical peatland. The southern section of this area is located in the north-east corner of Block A of the ex-MRP area; the northern section is located centrally within Block E of the ex-MRP area. In the past, this area was covered by peat swamp forest. However, the activities in ex-MRP projects bisected this area with a major canal and small hand-built canals that created different levels of land and forest disturbance. The extensive canals establishment occurred in Block A, as seen in Figure 7.1. The Block A area has been deforested and very severely degraded, while a large section of the Block E forest remains relatively undisturbed. In the KFCP area, fourteen community settlements are located along the rivers. These settlements formed nine village administrative units.

7.2 Test Data Preparation

This section covers the preparation of the dataset used for evaluating the causal model. The section starts with an explanation of the source of the hotspot data and the strategy employed to choose hotspots to be included as test data. This is followed by a detailed explanation on data for the causal variables associated with these hotspots collected from different sources and in various formats; how the climatic conditions data are obtained from measurements and estimations of satellite imagery, while the biophysical and human involvement data are gathered from government and non-government organisation (NGO) reports.

7.2.1 Source of Hotspot data

Daily Moderate Resolution Imaging Spectroradiometer (MODIS) hotspot data (Collection 5.1 active fire product) for the year 2012 have been extracted through the Fire Information for Resources Management System (FIRMS) website (http://earthdata.nasa.giv/data/near-real-time/data/firms). The temporal resolution of a single hotspot detected by MODIS satellites Terra/Aqua represents an area with a centre of approximately 1 km^2 per pixel (Giglio et al., 2016). There are 12 items included in the composite data, such as latitude, longitude, brightness, confidence, acquisition date, and acquisition time. In this research, not all the items in the dataset were considered in the analysis. Only latitude and longitude are used

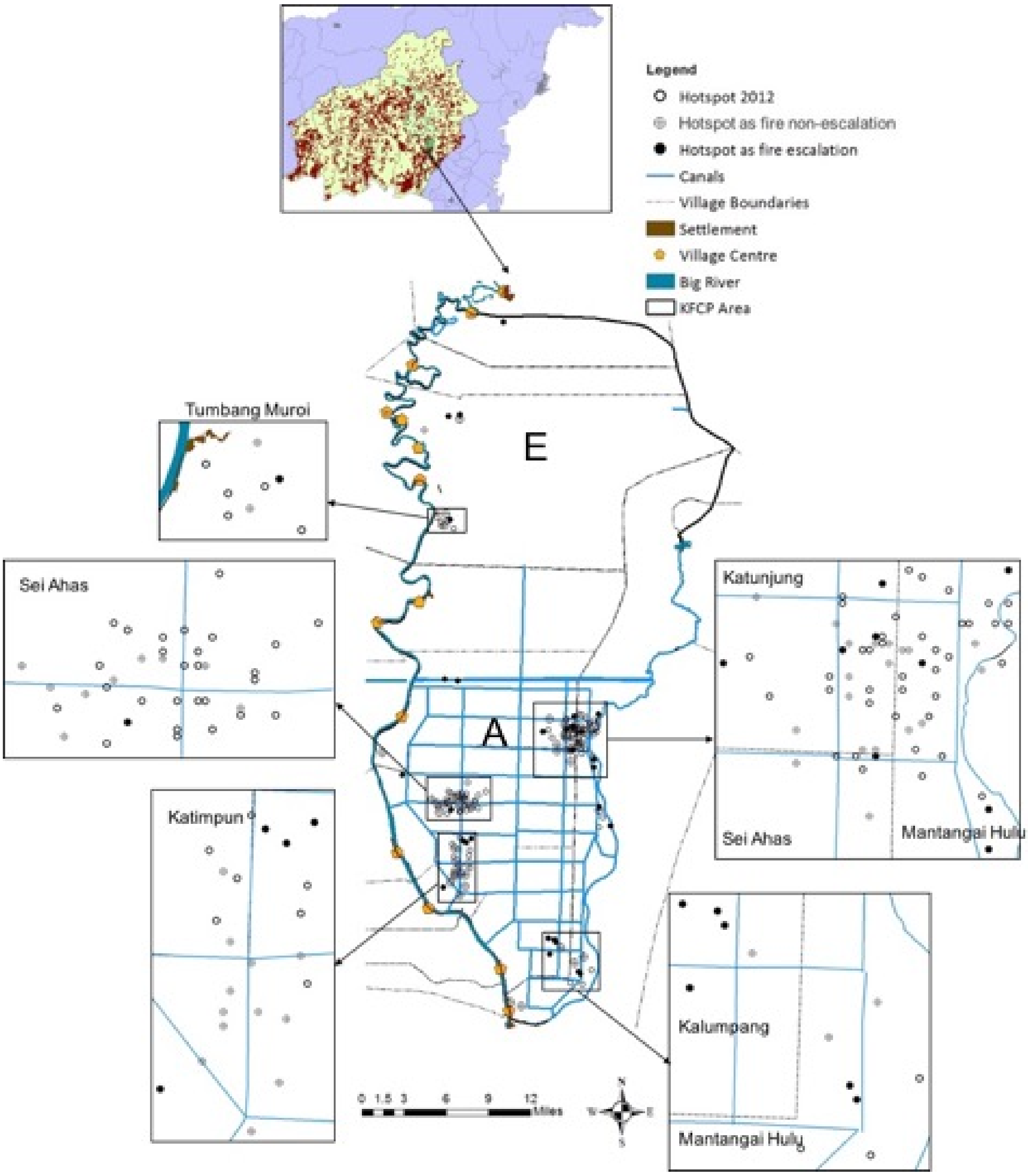


Figure 7.1: Study area, located in the KFCP project area. The white circles represent the hotspot occurrences in 2012. Black solid circles are hotspots detected as fire escalations by the KFCP team; circles with plus sign inside represent the hotspots as fire non-escalation. The four squares represent the clusters of the hotspots.

to represent the location of hotspots. The acquisition date and time are included in the data analysis.

In the year of 2012, the MODIS satellites detected 165 hotspots in the KFCP area. Most of the hotspots were located in Block A, as seen in Figure 7.1. In 2012, hotspots were detected mostly between the period of August and November 2012, during the dry season. 'Figure 7.2 shows that 95% of hotspots occurred in September and October 2012, which is the peak of the dry season in Indonesia. In this thesis, the hotspot data is used to evaluate the causal model.

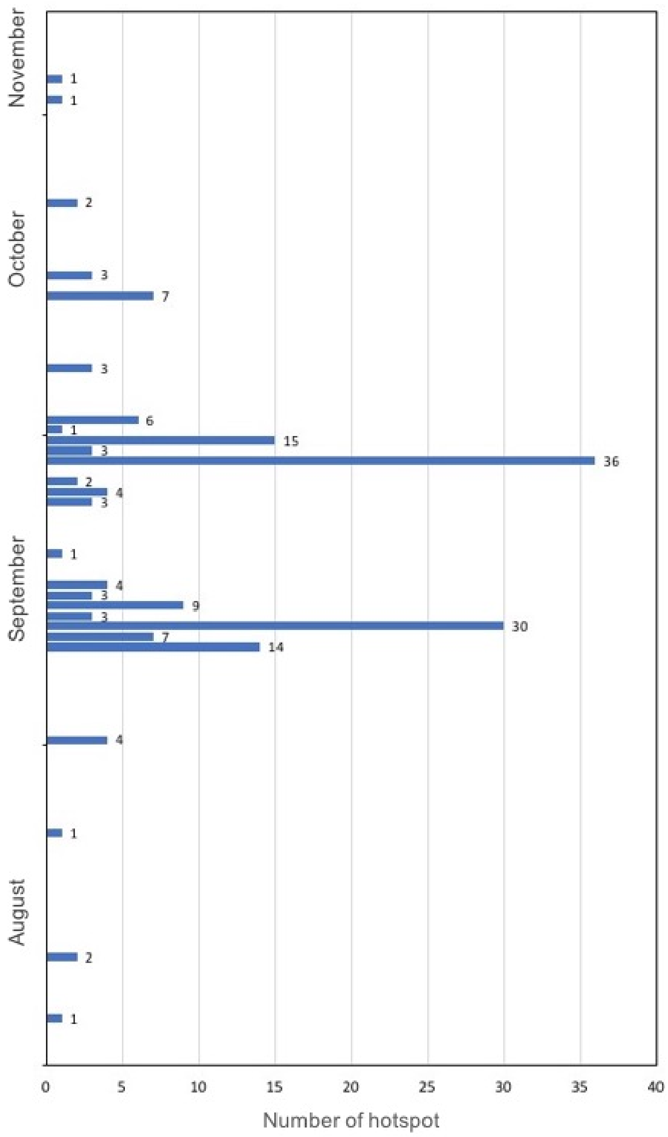


Figure 7.2: Hotspot distribution in the KFCP area from August - November 2012

7.2.2 Choice of Hotspot Data

Two categories of past hotspots are used for model evaluation. i.e. those that escalated and those that didn't escalate).

Data of Hotspot Escalations

As explained in Subsection 3.4.2, the fire escalation data has been obtained from the investigation of the KFCP Fire Management Team (FMTeam). Out of the 165 hotspots detected within the KFCP boundaries, the FMTeam have been able to investigate 30 hotspots and documented the hotspots as fire incidents and these are shown as the black dots in Figure 7.1 These hotspots are therefore considered ground truth data.

Data of Hotspot Non-Escalations

The 2012 ground fire investigation from the FMTeam did not include information on hotspots that did not escalate into wildfires. Therefore in this thesis, a number of assumptions are made to choose hotspots that could be categorised as non-escalations. A hotspot is assumed not to have escalated into a wildfire if:

1. It is a non-investigated hotspot. This means that the hotspot occurred on the same day as the ground checked hotspot, but was not investigated. Only the non-investigated hostpot that occurred outside the radius of 2km from the the ground check hotspot are considered non-hotspot escalations.

2. The hotspot occurred in the wet months such as July or late October to November.

3. It is a single hotspot that occured only for one day. For this criterion, we need to ensure that no cloud cover occurred on that day and the next day, since clouds might hinder hotspot detection. (Giglio et al., 2016).

Using these assumptions, 50 hotspots have been identified as fire non-escalations and they are indicated as circles with plus sign inside in Figure 7.1.

7.2.3 Data on Climatic Conditions

The climatic dataset used in this research has been obtained from the satellite estimation and analysis data. This dataset includes ENSO data, rainfall data, relative humidity data and wind speed data. The strength of ENSO events was determined based on Southern Oscillation Index method (SOI) index. The SOI index was gathered from http://www.cpc.ncep.noaa.gov/data/indices/. The SOI is negative during the warm El Niño phase and positive during the cold La Niña phase (Susilo et al., 2013) The rainfall data used in this thesis has been acquired from the daily Tropical Rainfall Measuring Mission (TRMM) precipitation data (TRMM 3B42 daily) derived from GES-DISC NOAA (https://mirador.gsfc.nasa.gov, while the relative humidity data has been acquired from the Earth System Research Laboratory (ESRL), Physical Sciences Division (PSD), ESRL Reforecast Version 2 Project of NOAA (http://esrl.noaa.gov/psd/forecasts/reforecast2/).

Estimations of rainfall and relative humidity using satellite data are used because weather data from the nearest weather station located in Kapuas is not available online. The next closest station from the study area is in Palangka Raya, located almost 100 km away. This long distance would influence the correctness of the local measurement.

Around the KFCP area, there are four TRMM pixels spatially distributed (see Figure 7.3). The Inverse Distance Weighting (IDW) is used to interpolate data from these points to estimate the rainfall and relative humidity values at a particular hotspot location at a point in time. This deterministic method seems suitable because the topography around the study area is mostly flat. Hence, it is not necessary to use elevation data as a secondary variable. IDW has been chosen because it is a relatively simple and efficient method to use. It requires relatively little input data and can be used in small scale area (Yang et al., 2015).

Using IDW, a weight is assigned to the TRMM value based on the distance to the interpolation location. The interpolated value is the weighted averages of the estimated values. The IDW formulas are given as Eq. 7.1 and Eq. 7.2

$$R_p = \sum_{i=1}^{N} w_i R_i \tag{7.1}$$

$$w_i = \frac{d_i^a}{\sum_{i=1}^{N} d_i^a} \tag{7.2}$$

where:
R_p: the unknown TRMM value in a hotspot location
R_i: the value of known TRMM pixel location
N: the number of TRMM pixels
w_i: the weighting of each TRMM pixels
d_i: the distance from each TRMM pixel to the location of hotspot
a: power, generally assumed as two as used in (Chen and Liu, 2012).

Daily rainfall and relative humidity were acquired for January 2012 - December 2012.

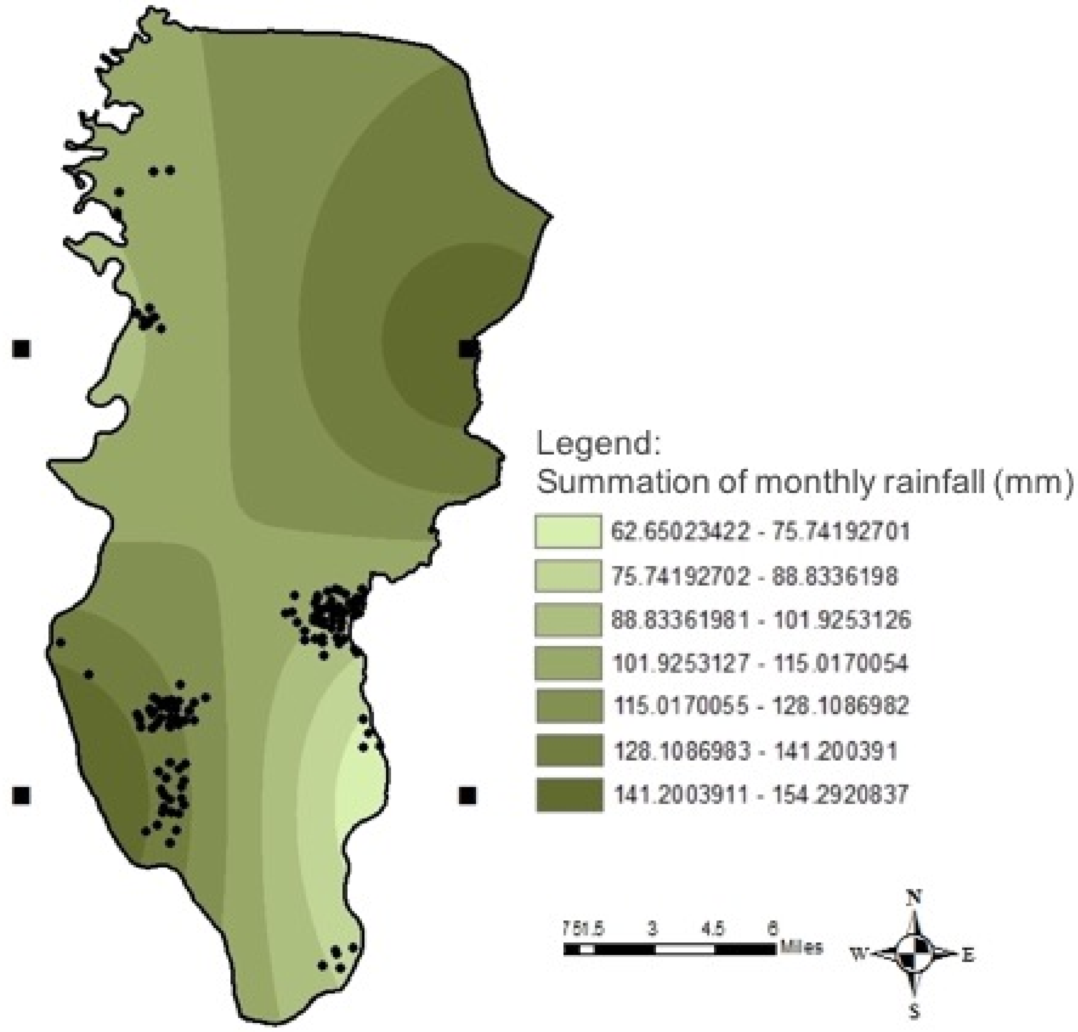


Figure 7.3: Estimated monthly rainfall for September 2012 in the KFCP area calculated using IDW interpolations. The black squares represent the four TRRM pixels which show the observed rainfall. The gradient colors in the study area represents the interpolation result using the IDW method. The black circles represent hotspot occurrences in September 2012

7.2.4 Human Factors

Data representing the human factors relevant to hotspot escalation is divided into two categories. The first category is information on the characteristics of the people living in every village in the KFCP study area. This information includes the livelihood of villagers, the size of the population, and the ethnicity of the people. This information has been gathered from Statistics Indonesia and non-government organisation reports.

The second category is the proximity of human access to the hotspots in the study area defined as the distance to the nearest settlement, the distance to the nearest road, and the distance to the nearest waterways. As shown in Figure 7.1, around the study area there are 14 village centres as part of nine villages. Locations of settlements have also been extracted from the land use and land cover map as additional information to the locations of village centres.

Waterways in the study area consist of two big rivers and small canals (see Figure 7.1). The locations of these waterways have been extracted from the KFCP database. The Kapuas river is located to the west and southwest of the study area, while the Mantangai river is to the east and southeast of the study area. Small canals are located mostly in Block A.

Planar measurement in 2D Cartesian coordinate system (available in ArcMap 10.5) has been used to get the distances from a hotspot to the nearest settlement and access route. In ArcMap 10.5, calculating distance is dependent on the geometry type of the features as well as other factors such as coordinate system. There are three basic rules to determine how distance is calculated:

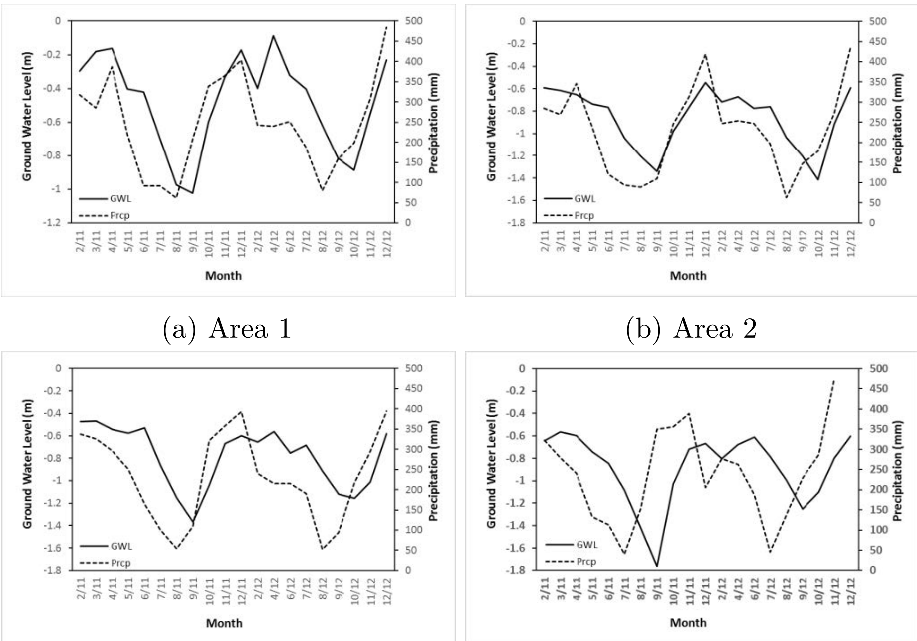
- The distance between two points is the straight line connecting the points. This rule is used to calculate the distance from a hotspot to the village centres.
- Distance from a point to a line is either the perpendicular or the closest vertex. This rule is used to calculate the distance from a hotspot to the settlement, the distance from a hotspot to the rivers and canals.
- Distance between polylines is determined by segment vertices. This rule has not been implemented in this research.

7.2.5 Biophysical Data

The biophysical data covers land use and land cover changes and the distribution of peat in the KFCP area. These data have been obtained from the KFCP database. The land use and land cover data map from the year 2010 was the latest documented land cover in that area. This has been used and overlaid throughout the area. Based on classification defined by the Ministry of Environment and Forestry, nine categories of land cover are found in the KFCP area: primary swamp forest, secondary dryland forest, secondary swamp forest, shrub-mixed dryland farm, bush/shrub, swamp, swamp shrub, settlement area, and water. The northern part of the area is still covered with primary forest, while the southern part is mainly an open area or non-forested area.

The peat decomposition level is divided into three categories: sapric, hemic and fibric. However, peat in the KFCP area only consists hemic and fibric, specifically 60% hemic and 40%fibric.

Another important relevant biophysical variable is ground water level. Although we are not able to source real measurement of ground water level for this project, experts have reported that ground water level and peat moisture content data could be derived from monthly rainfall data. There is a strong correlation between the amount of monthly rainfall with the fluctuation of ground water level. Putra et al. (2016) found that when the amount of rainfall gradually decreases, the ground water level also gradually decreases, starting from the month subsequent to the respective rainfall. As can be seen in Figure 7.4, Putra et al. (2016) measured the ground water level for four different locations in the KFCP area in 2011 and 2012. For example, it is shown that in 2012 when the rainfall gradually decreased in early July, the ground water level also gradually declined in early August. Therefore, in this research, monthly rainfall from the previous month is used in place of ground water level.



(a) Area 1 (b) Area 2

(c) Area 3 (d) Area 4

Figure 7.4: Ground water levels and rainfall patterns at four locations based on the TRMM pixels located around the KFCP Area. Source: Putra et al. (2016)

7.3 Test Results and Analysis

This section is divided into two subsections. The first subsection contains the analysis of the performances on the same test dataset, of the benchmark model, an implementation of the LAPAN's published guidelines of determining hotspot escalation (i.e. henceforth will be called the implementation of the LAPAN's rule) and of the causal model. The second subsection covers detailed analysis of the reasoning process of the causal model using using four different hotspot scenarios as illustrations.

7.3.1 Model Performances on Test Dataset

A set of performance metrics based on confusion matrices is applied to compare the performance of the implementation of the LAPAN's rule and that of the causal model.

7.3.1.1 LAPAN's rule

The LAPAN's rule, described in Chapter 3 Section 3.4.2, has been implemented and ran on test dataset, the result of which is presented in the confusion matrix in Table 7.1a.

The confusion matrix shows that the implementation of the LAPAN's rule has been able to correctly predict 19 hotspots as hotspot escalation into peatland fires. Eleven hotspots were incorrectly predicted as fire non-escalation. For the non-escalation hotspot dataset, the performance of the implementation of the LAPAN's rule shows there are 20 non-escalation hotspots incorrectly predicted as escalation and 30 hotspots correctly predicted as non-escalation hotspots. The accuracy of the implementation of the LAPAN's rule is slightly below that of the causal model (see Table 7.2), but it shows slightly better prediction for non-escalation hotspot dataset (see Table 7.1).

Using the implementation of the LAPAN's rule, more than 50% of the hotspots in the dataset have been correctly predicted. The result could be better if a comprehensive image processing in detecting smoke occurrence has been conducted. As described in Section 3.4.2, LAPAN uses RGB image processing to detect smoke occurrence. However, there is no detailed explanation on the procedure of the image processing algorithm to be used. Therefore in this thesis, a simple RGB processing (see Section 3.4.2) has been conducted. It is possible that the simple implementation from the RGB processing in this thesis might have produced in a worse model than the real implementation done by LAPAN.

7.3.1.2 Causal model of hotspot escalation into peatland fire

The result from the causal model using the available test data shows that the probability of a hotspot escalating into a peatland fire ranges from 44% up to 65% (see Figure 7.5). This figure shows that the distribution of probabilities from the

Actual \ Prediction	Escalation	Non-escalation	
Escalation	19	11	30
Non-escalation	20	30	50
	39	41	80

(a) Confusion matrix for LAPAN's algorithm

Actual \ Prediction	Escalation	Non-escalation	
Escalation	25	5	30
Non-escalation	25	25	50
	50	30	80

(b) Confusion matrix for the causal model result, T = 50%

Actual \ Prediction	Escalation	Non-escalation	
Escalation	23	7	30
Non-escalation	16	34	50
	39	41	80

(c) Confusion matrix for the causal model result, T = 51.1%

Table 7.1: Confusion matrix

causal model are right skewed. The causal model seems to result in a prediction that skews toward a hotspot escalating into peatland fire. The boxplot in Figure 7.6 confirms that the probability distribution of the causal model is asymmetric and right skewed. The median of the distribution cuts the box into two unequal pieces where the longer part of the box is above the median distribution.

The confusion matrix presented in Table 7.1b shows the performance of the causal model on the test dataset. Here, the causal model's probability of escalation threshold T = 50% is used, below this threshold, a hotspot is classified as a non-escalation. If the probability is above the threshold, the causal model predicts the hotspot as a peatland fire escalation. The detailed summary of the confusion matrix is:

- The number of hotspots correctly predicted as peatland fire escalation or *true positive* is 25

- The number of hotspots correctly predicted as non-peatland fire escalation or *true negative* is 25

- The number of hotspots incorrectly predicted as fire non-escalation or *false negative* is 5

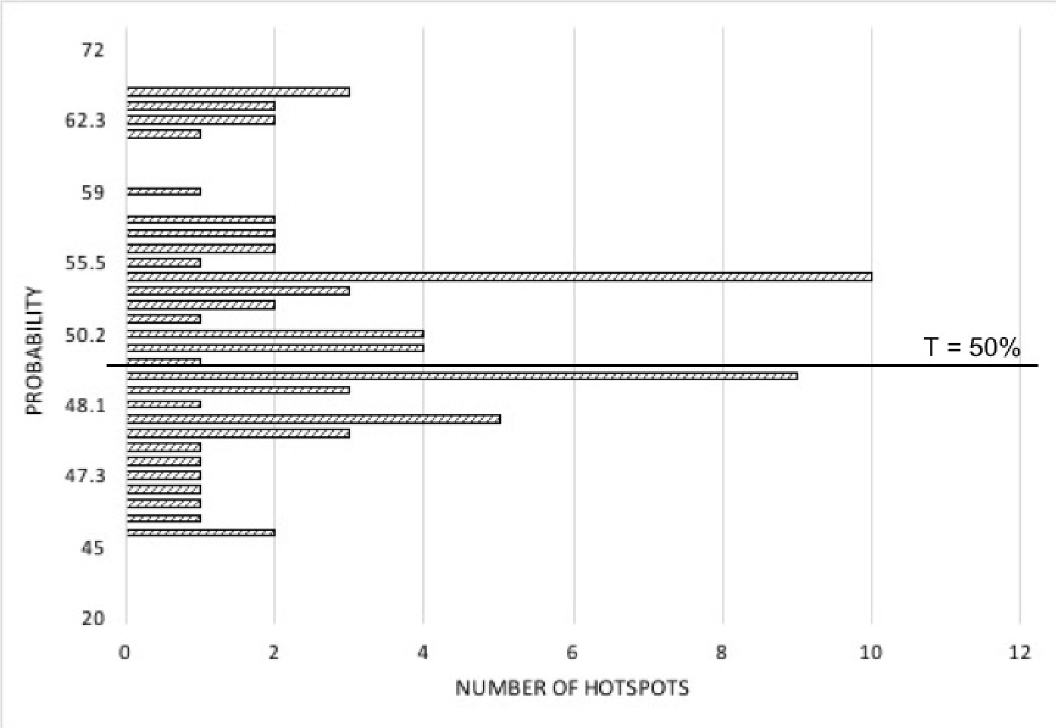


Figure 7.5: Probability distribution from the causal model result

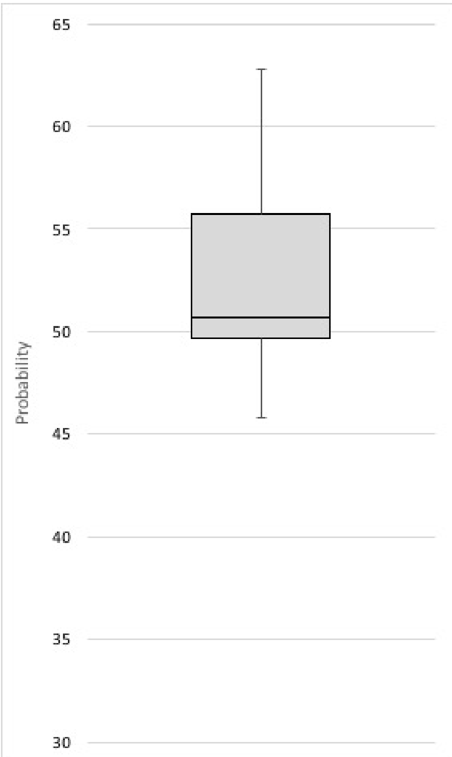


Figure 7.6: Boxplot of the probability distribution of the causal model result

- The number of hotspots incorrectly predicted as peatland fire escalation or *false positive* is 25

Table 7.2 shows other performance measurements on the performance of the causal model. For threshold T = 50%, the error rate is 37.5%. The model's ability to correctly predict which hotspot will escalate into peatland fire is represented by the sensitivity value, while the specificity value represents the model's ability to correctly predict hotspots as fire non-escalations. As shown in Table 7.2, the sensitivity is quite high (85%). This means that the causal model is relatively good in predicting hotspots that escalate into surface peatland fires. However the causal model has a much lower specificity of 50%.

Table 7.2: Model Assessment

Measurement	LAPAN's rule	Value for T = 50%	Value for T = 51.1%
Error Rate	38.75%	37.5%	28.75%
Accuracy	61.25%	62.5%	71.25%
Sensitivity	63%	83.3%	76.67%
Specificity	60%	50%	68%

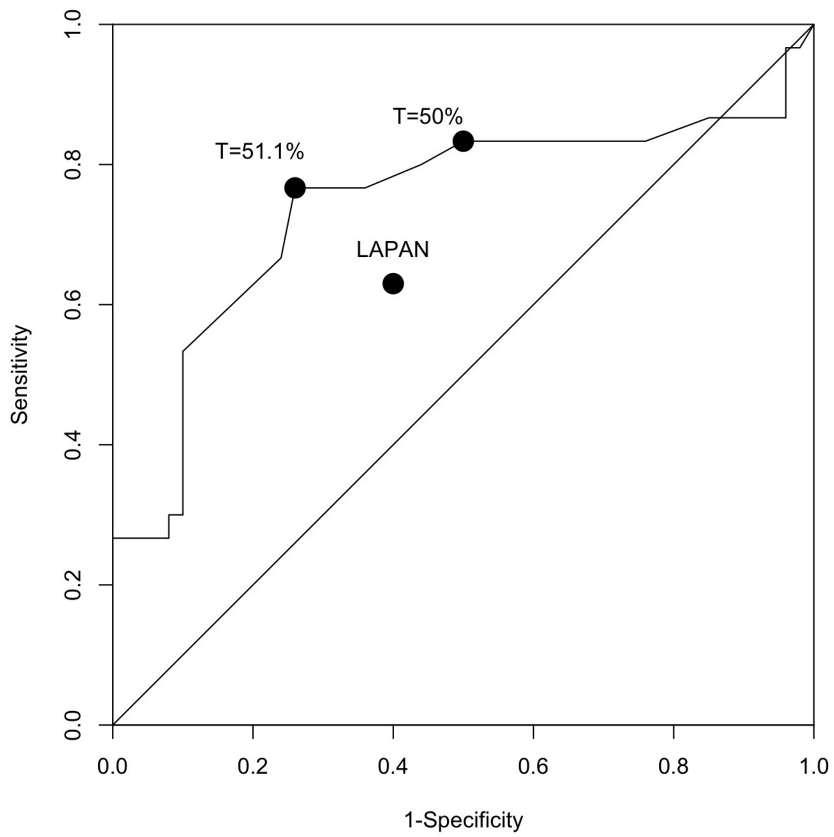


Figure 7.7: The receiver-operating characteristic (ROC) curve of the causal model

Based on the area under the ROC curve (AUC) of 0.746 shown in Figure 7.7, it can be concluded that the causal model demonstrates reasonable discrimination ability. The location of the model with T = 50%, on the ROC curve in Figure 7.7 is above the diagonal line. If the point is above the diagonal, it represents that the performance of the model is better than random guessing (Powers, 2011).

When a threshold of escalation probability of T=51.1% instead of T=50% is used for the causal model, as shown in Figure 7.7 the performance on non-escalation hotspot markly increases but the performance on escalation hotspot decreases slightly, i.e. as shown in Table 7.2, from 50% to 68% for specificity and from 83.3% down to 76.67% for sensitivity.

The purpose of developing a causal model for predicting fire escalation is as a decision support system. A wrong prediction will incur unnecessary economic and/or environmental costs. The choice of which threshold is to be used depends on which measurement is deemed more important for the domain. For the domain of peatland fire escalation in Indonesia, there is a perception that action would be taken only when the danger has become imminent. The actions of suppressing fire were more likely to be taken once there was visual indication of fire, such as smoke or wildfires (Saharjo, 2016). To encourage active fire mitigation, the Indonesian President, Joko Widodo, had issued a policy of firing the local military authorities if wildfires occurred in their territory (Ihsanuddin, 2018). Because of this it can be argued that a threshold of escalation probability of T=50% is to be chosen as it provides better sensitivity measure.

7.3.1.3 Sensitivity Analysis of Causal Variables: Information Gain Results

The sensitivity analysis was conducted on the belief of each node in the causal model based on the structure in Figure 7.8. The result of the sensitivity analysis of the causal model is summarised in Table 7.3. The variables/nodes are ranked according to the degree of their influence towards the outcomes of the fire escalation node, i.e. their reduction of entropy values (information gains).

As seen in Table 7.3, fire danger index is the most significant factor causing the largest entropy reduction. Peat flammability is the next most significant variable with near to 4% entropy reduction. An interesting result is found in the entropy reduction value of the 30-days cummulative rainfall. With an entropy value of close to 1%, this node is ranked above fire susceptibility and fuel flammability. This finding is consistent with the current findings in the literatures that ground water table is the a significant factor of fire escalation in peatlands. This is because the variable 30-days cummulative rainfall strongly correlates with ground water level.

Economic activities, activity on the land, and livelihood are the least influential variables to the peatland fire escalation nodes, with an entropy reduction value of zero. This result is contrary to what has been reported in the fire anthropology literatures on the unique characteristic of peatland fires in Indonesia. It is widely known that human activities in the peatland area are the most influential factors in the escalation of fires (Dennis et al., 2005; Vayda, 2010). This result suggests that the experts have been reluctant to assign decisively high or low probabilities to these variables, possibly because of the lack of certainty in the common understanding of the definition of these variables early on in the modelling process.

Table 7.3: Sensitivity Analysis of Causal Variables.

Node	**Entropy reduction**
Degree of fire index	0.03951 (3.9%)
Peat flammability	0.02177 (2.18%)
30-days Cum. Rainfall	0.00966 (0.967%)
Fire susceptibility	0.00798 (0.798%)
Fuel flammability	0.00521 (0.521%)
Peat moisture content	0.00506 (0.51%)
ENSO effect	0.00326 (0.326%)
Relative humidity	0.00298 (0.298%)
Rainfall from previous month	0.00110 (0.11 %)
Ground water level	0.00110 (0.11 %)
Dist. of hotspot from settlement	0.00049 (0.0469%)
Peat decomposition level	0.00006 (0.00586%)
Local wind speed	0.00005 (0.00544%)
Dist. of hotspot from access route	0.00003 (0.00263%)
Land condition	0.00001 (0.000818%)
Land use land cover change	0.00001 (0.000505%)
Population density	0.00000 (0.000151%)
Economic activities	0
Activity on the land	0
Livelihood	0

7.3.2 Scenario based Analysis of the Causal Model

Figure 7.8 shows the final version of the peatland fire escalation causal model including the probability distribution. There are 20 nodes representing the causal factors influencing the peatland fire escalation and 1 goal node, *surface fire escalation*. In this sensitivity analysis, the evaluation of the causal model is conducted based on the real hotspots of the study area in the year 2012 (described in Section 7.1).

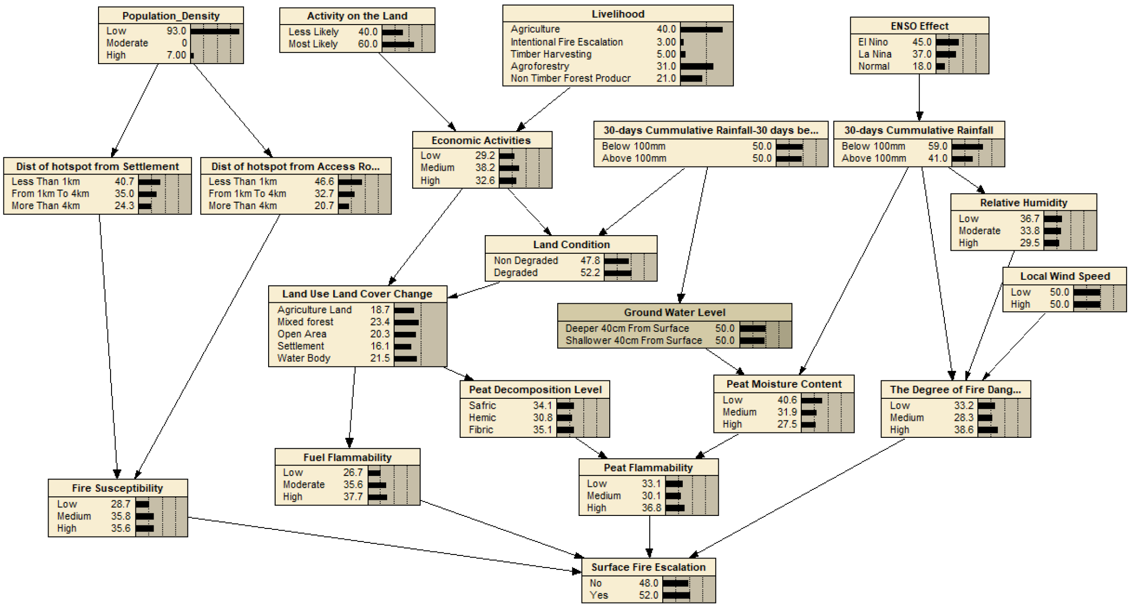


Figure 7.8: Causal model of peatland fire escalation before the constant values of node peat decomposition level and ENSO were included.

Based on the dataset provided by the KFCP FMTeam, one of the constant nodes in the causal model is *peat decomposition level*. The peat in the KFCP area only contains hemic and fibric decomposition levels, with proportions of 60% hemic and 40%fibric. Using the Oceanic Niño Index (ONI) to measure the El Nino-Southern Oscillation, 2012 could be categorised as a normal year. Therefore, the node ENSO also could be set as constant to state *normal*.

Before a few of the nodes in the causal mode were set into a constant condition, there was already a 52% chance that a hotspot would escalate into a peatland fire (see Figure 7.8). However, after the constant value in the node peat decomposition level and ENSO were included, the probability of escalation dropped to 50.9% (see Figure 7.9). Figure 2.1 taken from from Susilo et al. (2013) shows that during El Niño years, forest fires have a higher frequency compared to non El Niño year. Therefore in the causal model, when the value on *ENSO* node is set as state *normal*, the probability of hotspot will escalate into peatland fire is lower compared to condition of *ENSO* node in the state *El Niño*.

7.3.2.1 Analysis of Cases of Correct Escalation Predictions by theCausal Model

In order to explain the reasoning behind the correct predictions of the peatland fire escalation causal model, two scenarios are observed, one for ground-checked (ground truth) hotspot escalation and another for non-escalation.

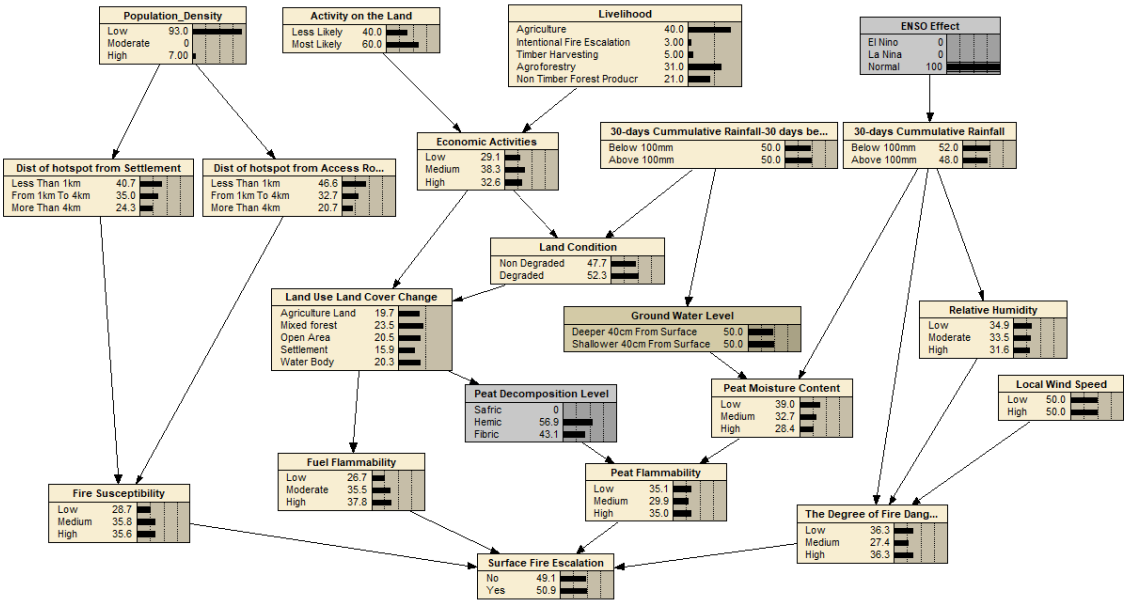


Figure 7.9: Causal model of peatland fire escalation after the constant values of node peat decomposition level and ENSO were included.

Scenario 1: Ground truth: fire escalation; Causal model: fire escalation

As shown in Table 7.4, both the implementation of the LAPAN's rule and the causal model correctly predicted eighteen hotspots as peatland fire escalation. The result from the causal model shows that all these hotspots have probabilities of fire escalation more than 50%, therefore it can be assumed that fire escalation was happening. These eight hotspots were also predicted as fire escalation by the implementation of the LAPAN's rule because the condition of these hotspots fulfilled the three rules presented in Subsection 3.4.2.

To be able to understand the reasoning of the result from this possible outcome, a scenario is generated based on the condition of the study area during the occurrence of a hotspot. This scenario is used as evidence and entered into the causal model.

- A hotspot occurred in Sei Ahas village (see Figure 7.10), one of the hotspots in Sei Ahas village, labelled as star. The main ethnic group living in this village is the Dayak people. About 5% of the population are people from the ethnic groups Banjar and Java. Logging timber is the most important source of livelihood in this village (44%), followed by agriculture such as paddy plantation (32%). There are small populations fishing and rubber tapping as their livelihood.

- The location of a hotspot is close to the canals, but the walking distance to the nearest settlement is outside a convenient radius.

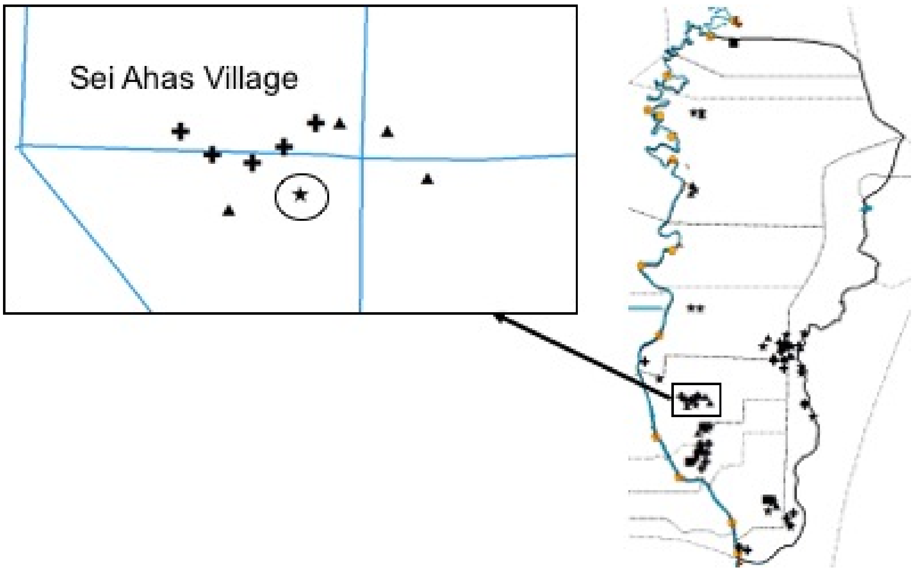
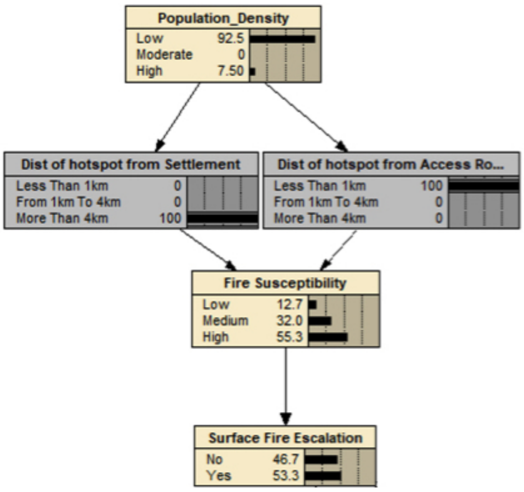


Figure 7.10: Locations of hotspot escalation and non-escalation in Sei Ahas village. The hotspot in the circle, labelled as star is correctly predicted as escalating into peatland wildfire.

- The weather condition during the hotspot occurrence was dry and windy. The summation of rainfall - 30 days before the hotspot occurred was 60mm/month, and it is below the threshold. The rainfall condition, one month before the hotspot occurred, was also below 100mm and also below the threshold. The relative humidity on the day was relatively low, and it was a windy day.
- The hotspot occurred in an area whose land use and land cover is agriculture.

As mentioned in the introduction of this section, before observing any scenario the probability of a hotspot escalating into a peatland fire is 50.9%. It means a nearly 51% chance that the hotspot will be escalating into peatland fire. If the location of a hotspot is observed less than 1 km or more than 4 km from the nearest settlement, the probability for node fire susceptibility rises to nearly 54%. This has raised the chance of peatland fire escalation to 53.3% (see Figure 7.11).

Even though the ENSO effect shows a normal condition, mid-September is categorised as a dry season. Therefore the amount of monthly rainfall could drop below the threshold of 100mm (Putra et al., 2011). If we also observe the climatic conditions when the rainfall was below 100mm, the probability of a high degree of fire danger rises to almost 52%, which also increases the chance of hotspot escalation to 58.9% (see Figure 7.12(a)). This result is supported by much research that found that due to the lack of rainfall in the dry season, the condition on the peatland area sometimes could be very dry and easy to catch fire (Usup et al., 2004). The combination of low rainfall with low humidity and windy conditions cause the degree of fire danger to rise to 58.3% and also slightly increases the chance of hotspot escalation above 60%



*

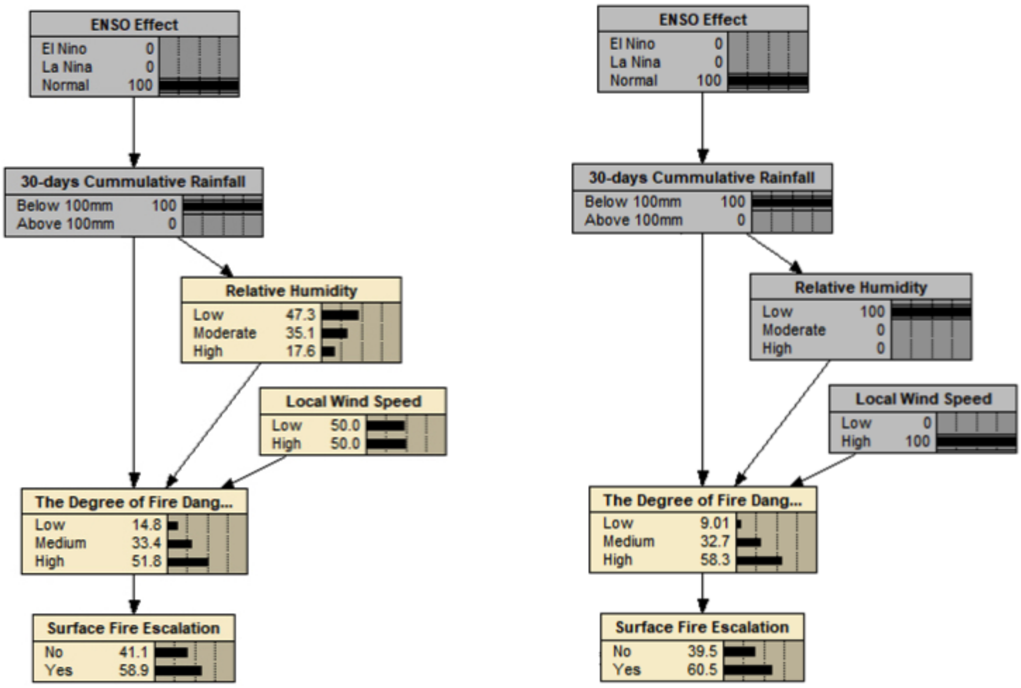
Figure 7.11: Predictive reasoning for the evidence of hotspot location to the nearest settlement and nearest access route

(see Figure 7.12(b)). This result shows that climatic conditions have quite significant influence on the hotspot escalation into peatland fires.

Another form of evidence, the rainfall conditions one month before a hotspot occurs, also has significant influence on fire escalation. Putra et al. (2016) found there is one month time-lag between rainfall and the ground water table. If the rainfall drops below 100mm/month, the ground water table in the next 30 days will also drop below threshold, which is deeper than 40cm from the surface. A lower ground water level is one of the favourable condition for the occurrence of peatland fire (Usup et al., 2004). When ground water level drops, it affects the dryness of the peat and the moisture content of the surface peat, making a suitable condition for fire escalation. In this scenario, the summation of one month of rainfall in the previous month is below 100mm. This condition has increased the chance of ground water level dipping below the threshold to 100%. Even though the evidence of peat decomposition level is a combination of hemic and fibric, the type of peat that would decrease the chance of hotspot escalation (as shown in see Figure 7.11), the poor condition of the ground water level doubles the chance of low peat moisture content. Therefore, the chance of a hotspot escalating into a peatland fire rises more than 2% from 60.5% (see Figure 7.12(a)) to 62.7% (see Figure 7.12(b)). This result shows that the causal model has been able to capture the influence of peat condition on hotspot escalation.

Scenario 2: Ground truth, non-escalation of hotspot; Causal model, non-escalation of hotspot

In this subsection, the analysis of the correct prediction for non-escalation of hotspots or *true negative* is presented. There are 10 hotspots correctly predicted as non-escalation of hotspot by the causal model and also the implementation of the LAPAN's rule. The hotspot was predicted as non-escalation, if the probability from



(a) (b)

Figure 7.12: Causal model of hotspot escalation into peatland fire: reasoning scenario - 1. (a) Predictive reasoning with rainfall above 100mm as evidence (b) Predictive reasoning with wind speed and relative humidity as additional evidence

the causal model was below 50%. While the implementation of the LAPAN's rule predicted a hotspot as non-escalation if the condition of that hotspot did not match any of the rules (as described in Section 3.4.2).

Predictive reasoning is conducted to explain the result from the causal model. A scenario is generated based on the condition of a non-escalation hotspot and used as input to the causal model.

- Hotspots occurred in Mantangai Hulu village (the green dots labelled as A and B in Figure 7.13). In this village, 98% of the population are Dayak people. Only 2% of people in the village had ethnicity of Banjarnese or Javanese. There is no foreign ethnicity living in this village. The most important source of livelihood is agroforestry, such as rubber tapper/plantation, followed by non-timber and forest products and agriculture.

- Both of the hotspots occurred less than 1 km from the canals. Hotspot A occurs about 1 km from the nearest settlement, while the hotspot B occurs a bit further from the settlement. These hotspots occurred within the walking distance of local people. See Figure 7.13

- This hotspot occurred on 16th October 2012. The weather conditions during hotspot occurrence were quite wet and windy. The summation of rainfall, 30 days before the hotspots occurred, reached 205mm/month. The rainfall condition, one month before the hotspots occurred, was just slightly above 100mm/month. Even though, the amount of rainfall was quite high, the relative humidity was relatively low, less than 50%.

- Hotspot A occurred in an agriculture area; Hotspot B occurred in land use and land cover as mixed forest.

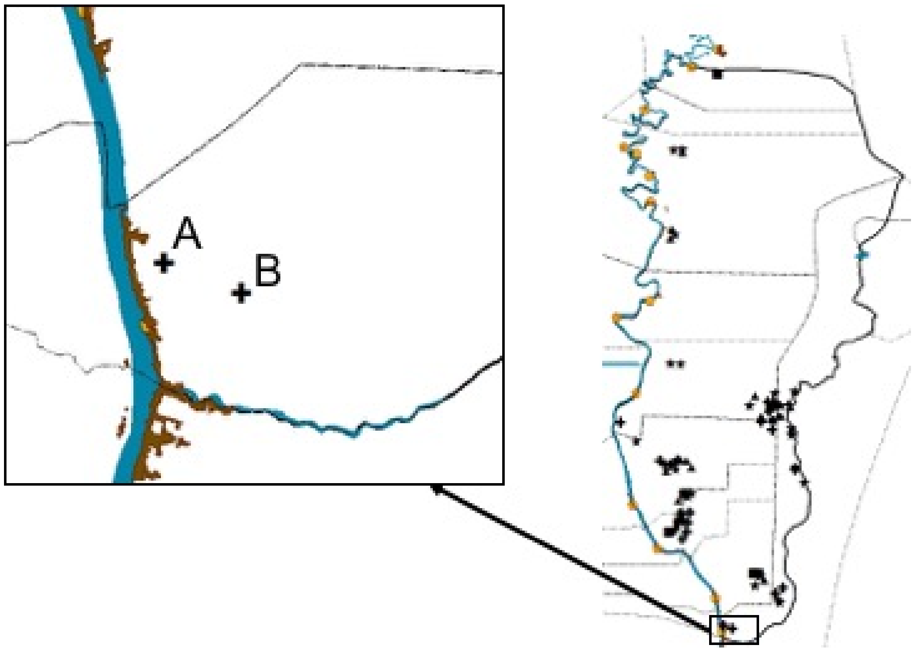
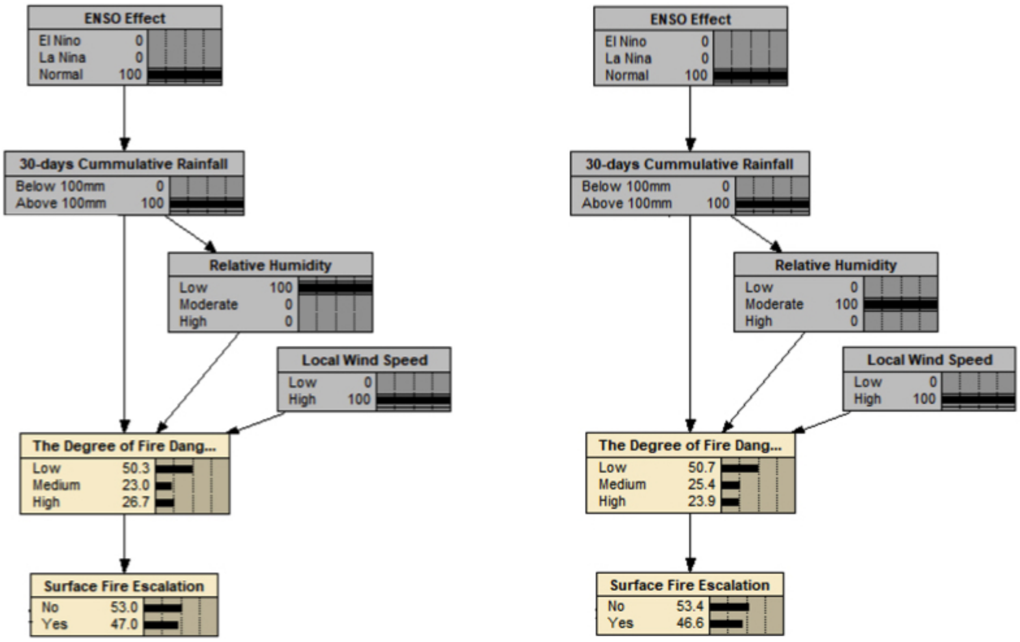


Figure 7.13: Location of non-escalation hostpots correctly predicted by the causal model of peatland fire escalation.

Based on the default input, the causal model shows there is nearly 51% chance of a hotspot escalating into a peatland fire (see Figure 7.9). Start the observation with the climatic condition category, since these two hotspots are have a similar value for this part. Based on the scenario above, the summation of rainfall was above 100mm. This high amount of rainfall has increased the chance of a low degree of fire danger to nearly 60%. As expected, when the degree of fire danger is low, the chance of hotspot escalation will drop. In this scenario, the probability of a hotspot is escalating into a peatland fire drops to 44.8% (see Figure 7.14(a)). The result of this scenario could be used to explain the finding in (Aiken, 2004) that was based on the investigation of fires in 1997. After heavy rains early in the wet season, all of the major fires in Sumatra and Kalimantan ceased abruptly.

When there is a high wind speed condition, the probability of a low degree of fire danger drops slightly and the chance of hotspot escalation also slightly increases. Although based on the experts' opinions for a context a rainfall above 100mm, there is a high probability of high relative humidity. The evidence from the scenario also shows low relative humidity. If the evidence of low relative humidity is entered into the causal model, the low degree of fire danger significantly drops to 50.7%. Thus, the chance of a hotspot escalating into peatland fire rises more than 2% (see Figure 7.14(b)). Also observed in this scenario, is a significant influence of rainfall on the degree of fire danger. Although the wind speed and relative humidity show a condition that might trigger a high degree of fire danger, the degree of fire danger tends to be low due to rainfall above threshold.



(a) (b)

Figure 7.14: Causal model of hotspot escalation into peatland fire: reasoning scenario 2 (a) Predictive reasoning with rainfall above 100mm as evidence (b) Predictive reasoning with wind speed and relative humidity as additional evidence

The next observation concerns human access. As shown in Figure 7.13, these two hotspots occurred less than 1 km from canals. In Indonesia, especially Central Kalimantan, fires tend to occur along the roads and canals (Liew et al., 1998; Yulianti et al., 2012). Therefore, the closer a hotspot is to an access route, the higher the chance of this hotspot escalating. As expected, when the distance from a hotspot to an access route is small, the chance of the hotspot escalating into a peatland fire is slightly increased by 0.3% (Figure 7.15(c)). The change is not really significant, but it shows that the causal model takes into account the influence of the access route on the escalation of a hotspot into a peatland fire. The distance from a hotspot to the nearest settlement is also observed. For Hotspot A, the distance is set to less than 1km, while for Hotspot B, the distance is set to between 1 km and 4 km. The result for this evidence can be seen in Figures 7.15(a) and 7.15(b). Even though there is a low chance that these two hotspots will escalate into peatland fire, there is a slightly difference in the probability value. Hotspot A has slightly lower chance of escalating into a wildfire compared to Hotspot B. The location of Hotspot A, closer to the settlement compared to Hotspot B, has influenced the likelihood of fire susceptibility. As seen in Figure 7.15(a), the chance of low fire susceptibility for Hotspot A reaches 34.4%. When the distance of Hotspot B to the nearest settlement is entered as evidence, the chance of low fire susceptibility drops to 20.9%. The chance of a hotspot escalating into a peatland fire rises to 47.3% (see Figure 7.15(b)). When a hotspot occurs within walking distance of a community, it is less likely to be escalated. This short distance has made it easier for the villagers to reach the fire location and prevent it from spreading to a wide area (Sumarga, 2017).

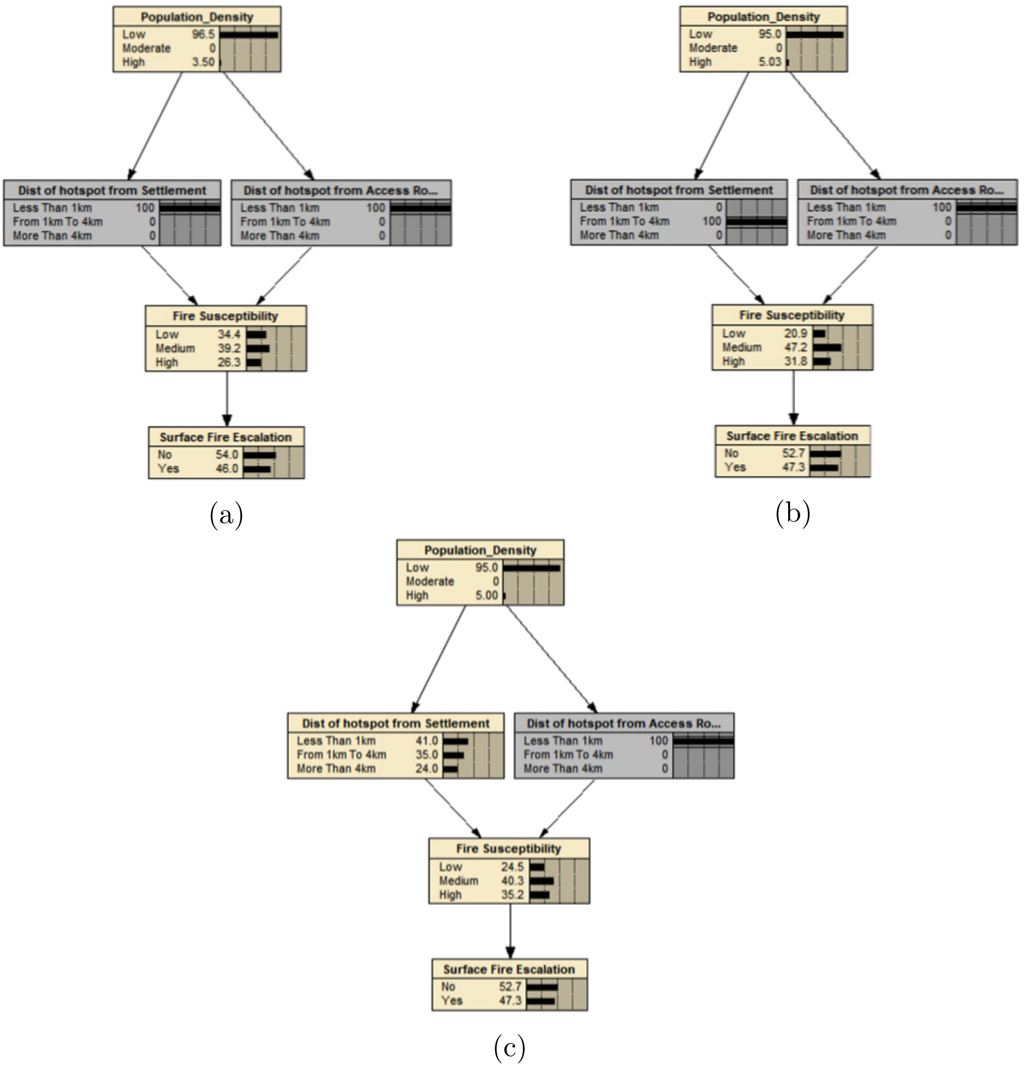


Figure 7.15: Causal model of hotspot escalation into peatland fire: reasoning scenario 2 (a) Predictive reasoning with a hotspot to settlement distance less than 1km as evidence. (b) Predictive reasoning with a hotspot to settlement distance between 1 km - 4 km as evidence. (c) Predictive reasoning for hotspot to access route distance only.

7.3.2.2 Analysis of the Incorrect Prediction of Peatland Fire Escalation Causal Model

In order to explain the reasoning behind the in-correct predictions (false positives and false negatives) of the peatland fire escalation causal model, two scenarios are observed, one for ground-checked (ground truth) hotspot escalation and another for non-escalation.

Scenario 3: Ground truth, hotspot escalation; Causal model, non-escalation hotspot

As shown in Table 7.4, both the implementation of the LAPAN's rule and the causal model incorrectly predicted four hotspots as non escalation hotspot. The result from the causal model shows that all these hotspots have a lower than 50% probability of hotspot escalation, therefore it can be assumed that no hotspot escalation was happening. These four hotspots were also predicted as non escalation hotspot by the implementation of the LAPAN's rule because the condition of these hotspots did not fulfilled the three rules presented in Subsection 3.4.2.

There is one hotspot in the fire escalation dataset that could be considered as an anomaly or outlier. This hotspot was located on the northside of Block E, occurring as single hotspot (see Figure 7.1). Based on the FMTeam guidelines in Subsection 3.4.2, this hotspot most likely will not be investigated as a fire escalation. However, since this hotspot occurred around the community's assets (Graham et al., 2014), the ground observation was conducted immediately and found as peatland fires after the investigation.

In order to conduct the diagnostic reasoning of this incorrect prediction, a hotspot occurrence scenario is generated based on the conditions of the day. This scenario is used as evidence and entered into the causal model.

- The hotspot occurred in Petak Puti village. The main ethnic group in this village is Dayak. There is a small percentage of Banjarnese and Javanese also living this village. Fishing is the main source of livelihood (58%) followed by agriculture such as rubber plantation as the second most important livelihood in this village.

- The location of the hotspot was quite close to the nearest settlement, less than 2 km. The hotspot occurred almost 3 km from the nearest river.

- This hotspot occurred on the 5th of August 2012. The weather conditions during this hotspot occurrence were quite wet and windy. In early August 2012, the summation of 30 days rainfall was quite high, almost reaching 200mm/month.

> Early August is the beginning of dry season. The summation of 30 days rainfall one month before the hotspot occurrence was also high, almost reaching 150mm. Both of these rainfall conditions were above the threshold of the dry season. The rainfall still occurred quite often around this time (see Figure 7.4).

The observation of default evidence in the causal model is that the probability of hotspots escalating into peatland fires is nearly 51% (see Figure 7.9). After the evidence in this scenario is entered, the probability of hotspots escalating into peatland fires drops to 45.8% (see Figure 7.16). This result means there is a low chance that the hotspot will escalate into peatland fires. This contradicts the result of the ground investigation from the FMTeam.

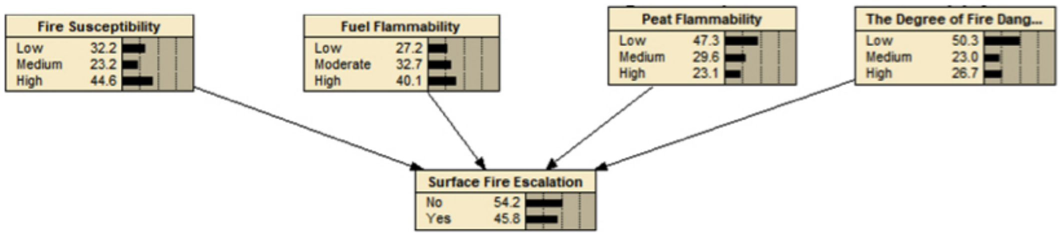


Figure 7.16: Result from the causal model based on the evidence in scenario 3 .

When the surface fire escalation shows high probability in the state *No*, it is expected that the influencing nodes (*fire susceptibility*, *fuel flammability*, *peat flammability* and *degree of fire danger*) should have high probability in poor conditions. Figure 7.16 shows the probability of peat flammability in state *low* and degree of fire danger in state *low* are highly probable, almost double the probability in the other states (*medium* and *high*). These results are also expected, as mentioned in the scenario, both the summation of rainfall 30 days before the hotspot occurrence and also the summation of rainfall in 30 days of the previous month are above the threshold. Once the amount of rainfall is above the threshold, the degree of fire danger definitely will be low. This is also applied to peat flammability.

However the node *fuel flammability* shows opposite result. The chance of *fuel flammability* being high is about 40%, higher than the other states. This is actually influenced by the fact that the hotspot occurred around the agriculture area. Once this evidence is entered into the causal model, the probability of high fuel flammability rises from 37.8% to 40% (see Figure 7.17). However, the probability distribution of each state combination in this node is not really diverse. As it can be seen in Figure 7.17, there is no significant diversity in the probability between the states low, moderate, and high. This could be triggered by the experts being indecisive in giving the probability (as explained in Chapter 6). Even though it is commonly known that the activities in the agricultural areas, such as land clearing, have significant influence on fire escalation (Dennis et al., 2005; Harrison et al., 2009).

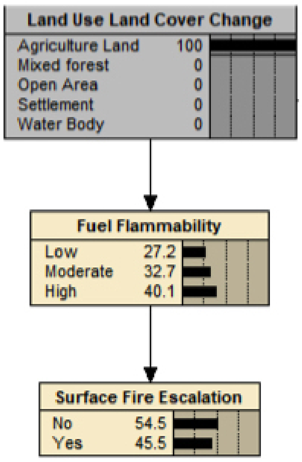
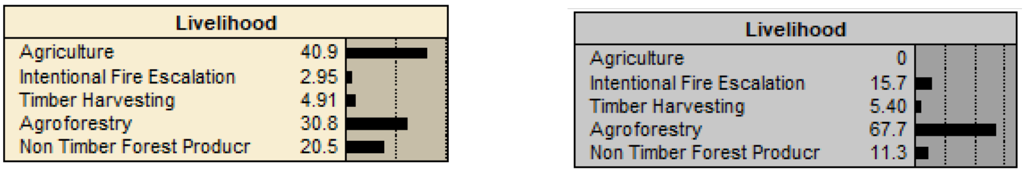


Figure 7.17: Result from the causal model based on the evidence provided in scenario 3.

Based on the investigation from the KFCP FMTeam, there are no records on the cause or motivation of this fire. The ground investigation only found that the fire location was close to the access route. Based on this location, there is possibility it is an accidental fire or an abandoned fire. As mentioned in the scenario, the main source of livelihood in this village is fishing. There should be a few fishing locations or attractions located around this area. Therefore, there is a high chance that people came to the area for the purpose of fishing, initiated the fires, and just left. The causal model also considers fishing activities as one of the highly influential factors in hotspot escalation under the node *livelihood* with state *intentional fire escalation*. However the default marginal probability that is assigned for the intentional fire escalation is quite low. There is only a 3% chance of this state influencing the escalation of a hotspot (see Figure 7.18(a)). As a result, when evidence was entered into the node *livelihood*, with the state *intentional fire escalation* given a high chance, the marginal probabilities assigned for this state are still low, much lower than the *agroforestry* which is the second means of livelihood (see Figure 7.18(b)). It has been investigated that fishing activities in some areas in Indonesia, especially in Cental Kalimantan, involve fires (Vayda, 1999; Tacconi, 2003; Chokkalingam et al., 2005). People use fire to burn the vegetation surrounding the fishing spots. These fires, when uncontrolled and expanded could be the major factor in widespread fires (Chokkalingam et al., 2005). Based on this observation, reassigning the marginal probability for the states in the node *livelihood* might be an option to capture the phenomena of fishing activities.

There is a possibility that the wildfires escalated from land clearing activities that used fire. Based on the report from the ground checked, one-third of fires were found on the land that belong to villagers. The causal model does not capture information about the ownership of the land as a factor that contributes to hotspot escalation into



(a) (b)

Figure 7.18: Causal model of hotspot escalation into peatland fire: reasoning scenario 3 (a) Default marginal probability for livelihood node (b) Marginal probability after the evidence was entered

peatland fire. The relationship between land ownership and fire escalation was briefly discussed during the evaluation workshop (see Chapter 6). Since the discussion was about the influence of planned burning on fire escalation, using the term *activity on land* was more suitable than the term *land ownership*. Based on the observation of this scenario, adding information about land ownership into the causal model might be useful when considering causal factors.

Scenario 4: Ground truth: non-escalation of hotspot ; Causal model: hotspot escalation

The causal model also incorrectly predicts a few non-escalation hotspots as hotspots escalating into peatland fires.

In order to conduct the diagnostic reasoning of this incorrect prediction, a hotspot occurrence scenario is generated based on the conditions on the day. This scenario is used as evidence and entered into the causal model.

- A hotspot occurred in Sei Ahas village (see Figure 7.19, a hotspot in a circle, labelled as triangle). The main ethic group living in this village is the Dayak people. About 5% of the population are people from ethnic groups Banjar and Java. Logging timber is the main source of livelihood in this village (44%), followed by agriculture such as paddy plantation (32%). There are small proportions of people fishing and rubber tapping.

- The location of hotspot 4.2km from the nearest settlement and less than 0.5 km from the access route.

- This hotspot occurred on 28th of September 2012. Nearly to the end of September 2012, rainfall started occurring more frequently. The summation of 30 days rainfall was above 100mm/month. The summation of 30 days rainfall-one month before the hotspot occurrence was quite low, only 49mm/month. This amount of rainfall was below the threshold, which might influence the ground water level.

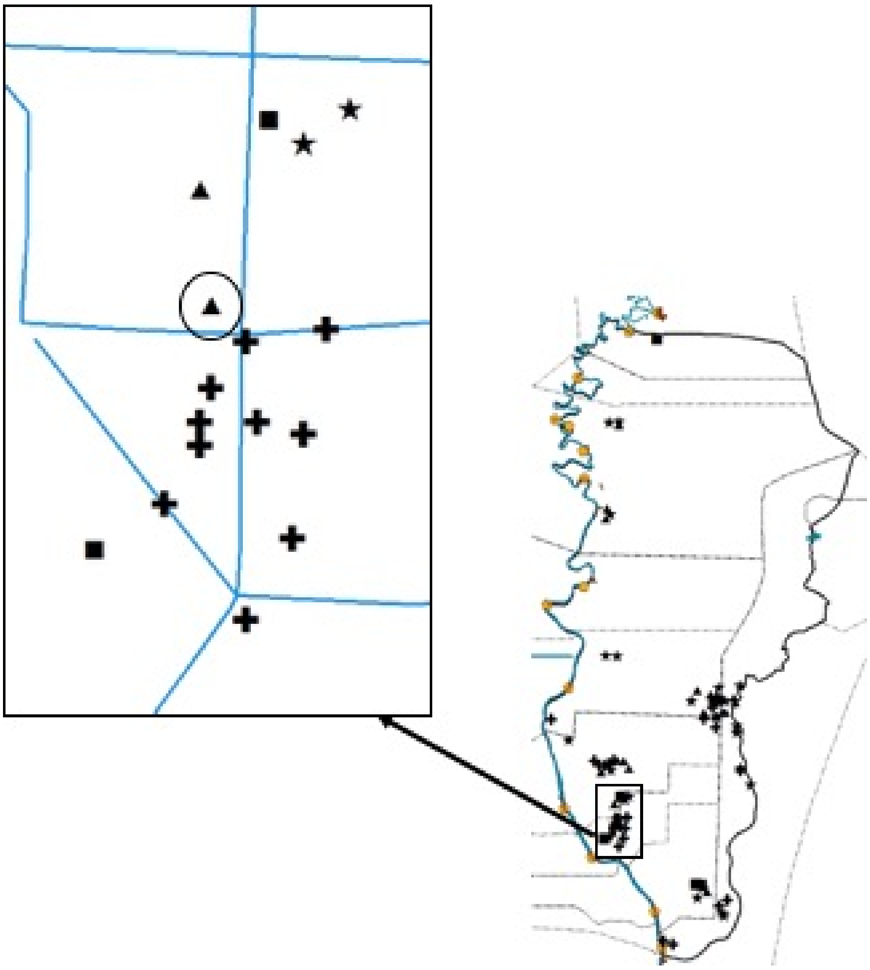


Figure 7.19: Locations of hotspot escalation and non-escalation in Katimpun village. The hotspot in the circle, labelled as triangle is incorrectly predicted as escalating into peatland wildfire.

As shown in Figure 7.20, the result from causal model shows there is a chance of 50.1% that the hotspot is escalating into peatland fire, even though the ground checked data confirmed that this hotspot is a non-escalation of hotspot. The result from the causal model was most likely influenced by the high probability in state *high* of three parent nodes, fire susceptibility, fuel flammability and peat flammability. Only node *degree of fire danger* had a high chance of being in a low state. Evidence influencing the probability of those parents node is investigated. In the node *degree of fire danger*, there is 55% probability of low fire danger. This is because the summation of 30-days of rainfall before the hotspot occurred was above 100mm. In addition to this rainfall condition, on the day the conditions were not windy. Therefore, it makes sense that the degree of fire danger was low.

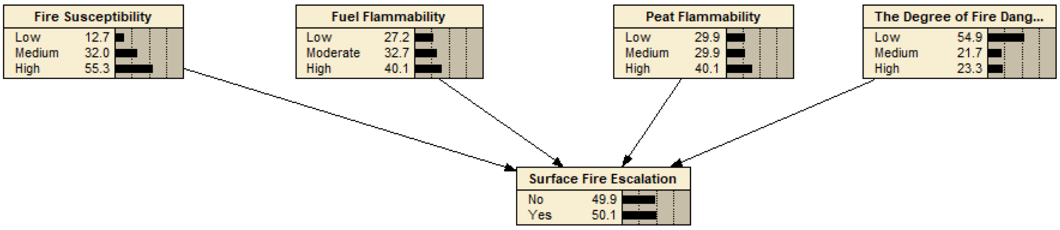


Figure 7.20: Result from the causal model based on the evidence provided in Scenario 4.

The low amount of rainfall in the previous month has raised the chance of the ground water level becoming deeper 40cm from the surface up to 100% (see Figure 7.21). The high amount of current rainfall might make the surface of peatland become moist and wet, but the chance of low peat moisture content is still high. This happens due to the ground water level being given more weight in the compatible parental configuration. Research has found that ground water level has significant influence on the condition of peat moisture content and the flammability of the peat (Usup et al., 2004; Putra et al., 2011).

Fire susceptibility also shows a high chance of influencing hotspot escalation. This because hotspot occurred about 4.2 km from the nearest settlement. It is only 200m further away from the 4km threshold of walking distance. But the problem is in this causal model. Once the distance of the hotspot to the nearest settlement is more than 4km, it is automatically placed under the third state. It will generate the same result as a distance of hotspot more than 10km. However, if the measurement is adjusted, the result might change. An example is rounding down/up the distance to the nearest integer value. It means the distance of 4.2 km could be considered as 4 km. The condition of state will fit in the second state, and reduce the high chance of fire susceptibility (see Figure 7.22(b)). Based on the report from Steenis and Fogarty (2001), the settlement was found more than 5 km from fire ignition zone. If this result investigation result is taken into account and the states adjusted in the node,

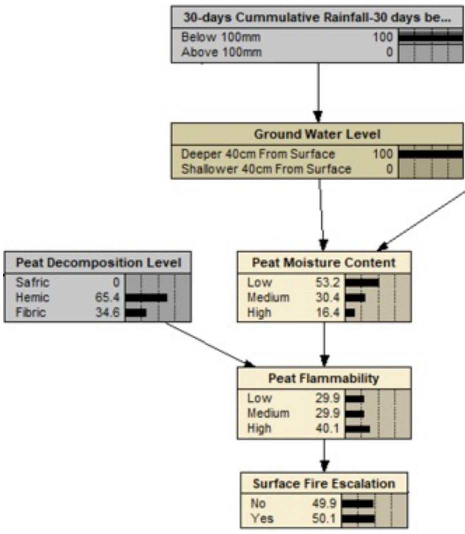
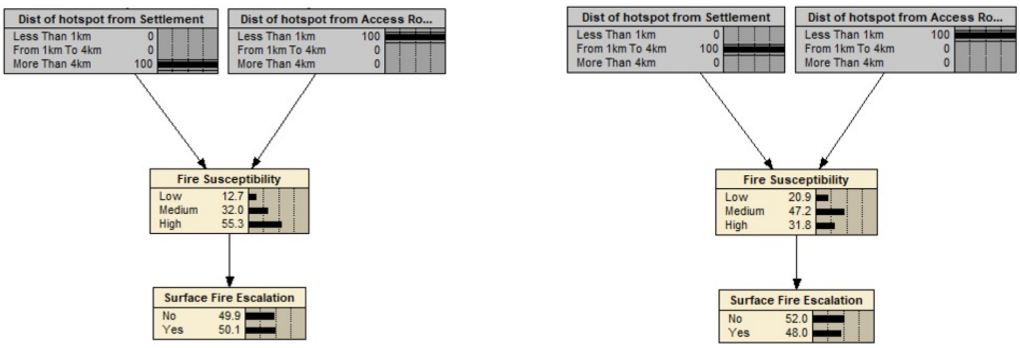


Figure 7.21: Result from the causal model based on the evidence provided in scenario 4.

the result might be change. However, their finding was based on the investigation in West Kalimantan. Whether it is applicable for the peatland fire situation in Central Kalimantan, further investigation should be done.



(a) (b)

Figure 7.22: Probability resulting from the causal model of hotspot escalation into peatland fire: reasoning Scenario 4 (a) The condition if a distance of 4.2 km is placed in the third state (b) The condition if a distance of 4.2 km is placed in the second state

7.4 Comparison of the Causal Model Result with the Implementation of the LAPAN's Guideline

As explained in Subsection 3.4.2, the result from the causal model of hotspot escalation with the prediction using a guideline provided by LAPAN is compared against the ground truth data from the KFCP database. The result from the implementation of the LAPAN's rule and the causal model are already described in Subsection 7.3.1. In this section, the comparison is presented in order to measure the performance of each model against the ground truth data. Eight possible outcomes are presented in the result of the comparison in Table 7.4. For the detection of the hotspot escalation into peatland fires, the causal model delivers better prediction. Out of 30 hotspot points in the fire escalation dataset, the causal model has been able to correctly predict 25 hotspots or 83% of the fire escalation dataset, while the implementation of the LAPAN's rule is able to correctly predict 18 hotspots, or almost two-thirds of the dataset. Based on this result it can be said that the causal model provides better prediction compared to LAPAN's rule in predicting hotspot escalation. However, during the implementation of the LAPAN's rule, one of the rule's might not have been implemented as accurately as LAPAN. The RGB image processing applied in this research might not be as comprehensive as implemented by LAPAN.

Table 7.4: Comparison between the result from the causal model of peatland fire escalation and the implementation of the LAPAN's rule in regards to ground truth data.

<table>
<tr><th>Ground Truth Data</th><th colspan="4">Escalation (E)

30 hotspots</th><th colspan="4">Non-escalation (NE)

50 hotspots</th></tr>
<tr><td>**Causal Model**</td><td colspan="2">Escalation
25</td><td colspan="2">Non-escalation
5</td><td colspan="2">Escalation
25</td><td colspan="2">Non-escalation
25</td></tr>
<tr><td>**LAPAN**</td><td>**E**
18</td><td>**NE**
7</td><td>**E**
1</td><td>**NE**
4</td><td>**E**
10</td><td>**NE**
15</td><td>**E**
15</td><td>**NE**
10</td></tr>
</table>

7.5 Chapter Summary

In this chapter, the evaluation of the performance of the causal model in predicting the hotspot escalation and non-escalation hotspot is presented. The performance evaluation was conducted based on the error rate and accuracy of the causal model in predicting the escalation of a hotspot into a peatland fire. Based on the error rate and accuracy result, the causal model incorrectly predicted almost one-third of the dataset. However, the sensitivity result shows that the causal model delivers good performance in predicting the hotspot escalation into peatland fires. The threshold T=50% also was chosen to determine the probability of hotspot escalation and non-escalation hotspots. By using this threshold, the causal model delivers a better prediction of hotspot escalation into peatland fire.

The sensitivity analysis was also conducted in order to ensure the causal model captured the characteristic of peatland fires. Based on four scenarios generated for the analysis, it was shown that the causal model has been able to show *correct* reasoning in the climatic conditions and peat ecology. The causal model was also able to quantify the human involvement influences on the escalation of hotspots. The reasoning was mostly presented in the narrative form. However, the causal model also incorrectly predicted hotspot escalation and non-escalation. Overlooked influencing factors and parameters, the indecisive experts in providing the probability for the causal model have become the factors of this inaccurate prediction.

The performance of the causal model was also compared with the result from implementation of the LAPAN's rule. The guideline of predicting fires issued by LAPAN was implemented and evaluated using the hotspot escalation and non-escalation dataset. The result from the causal model is slightly better compared to LAPAN's rule. However, the rules about smoke occurrence might not be implemented in the same way as LAPAN. Due to there being no clear explanation of how LAPAN detects smoke occurrence, the RGB image processing of smoke detection conducted in this research might not as deep or comprehensive as LAPAN's image processing. Therefore, the result from the implementation of the LAPAN's rules might improve if comprehensive RGB image processing is conducted.

Chapter 8

Conclusion and Future Work

8.1 Conclusion

This thesis contains an explanation of cross-disciplinary research that involves peatland fire science and data science technology, along with elements of complex human behaviour. A workflow for the development a causal model of hotspot escalation into peatland fires using the experts' knowledge is proposed. The conclusions of the research are discussed in this chapter.

8.1.1 Thesis Summary

The frequent occurrence of peatland wildfires in the province of Central Kalimantan, Indonesia is a damaging environmental problem on a global scale. These fires have not only destroyed a million hectares of Indonesian forest, but have also produced haze and released carbon into the athmosphere, causing economic and health problems, and contributing to global greenhouse gas emission problems. As has been advocated by various forest fire experts, to alleviate the occurrence of peatland wildfires it is important to ensure that hotspots during the dry seasons do not escalate into wildfires in the first place. Having a better understanding of the behaviour of peatland fires could provide more reliable information that can enable better decision making when predicting the escalation of hotspots into peatland wildfires (Applegate et al., 2002; Dennis et al., 2005).

This thesis proposes a workflow to develop a causal model that can explain the behaviour of surface peatland fires in Central Kalimantan. The model is used to predict which hotspots may escalate into peatland fires. Research in modelling forest fires in prediction systems mostly depends on historical data. However, for peatland fires in Central Kalimantan, insufficient historical data became a challenge to build a data-driven model of hotspot escalation into peatland fire. Therefore, the workflow incorporates information from literature and knowledge from experts to develop the causal model. The model is developed in the form of probabilistic graphical model using Bayesian Networks to incorporate all the factors influencing the behaviour of peatland fire.

The development of the causal model is started with identification of causal variables of hotspot escalation into peatland fire. Chapter 4 demonstrates how text mining analysis using topic modeling can be used to extract the causal variables of peatland fire from the literature. The result of this process indicates that in the situation where no domain experts are available to provide the causal variables for a causal model, automation of information extraction from the literature may be used to replace the domain experts' contributions. These causal variables then are used in both the initial construction and subsequent further development of the structure of the causal model, as explained in Chapter 5. A focus group discussion with fire experts was conducted to build and refine the structure of the causal model. A few iterations of evaluation using the Bayesian Network guidelines were conducted to refine the causal model structure. Once the structure of the causal model was developed, an online survey was setup to gather the probability distribution. In Chapter 6, an implementation of the Das method (Das, 2004) was explained. This method was used to generate the questions for an online survey and populate the conditional probability table (CPT). Chapter 7 presents the evaluation of the performance of the causal model in predicting hotspot escalation and non-escalation hotspots. The analysis of the evaluation result shows that the causal model has been able to show correct reasoning in the climatic conditions and peat ecology. The causal model was also able to quantify the human involvement influences on the escalation of hotspots.

8.1.2 Addressing the Research Questions

The introductory chapter outlines the research questions for this thesis:

RQ.1: What can be learned from current approaches used in prediction models for escalation of peatland fires in Indonesia?

This research question is answered in great detail in Chapter 2, specifically Section 2.2. In this Section the use of current approaches in predicting the fire occurrences used by the Indonesia fire authorities is explored. In addition, an exploration of the approach around the prevention system for forest fire occurrence is also presented. This question has two sub-questions:

RQ.1.1: What are the current approaches in predicting peatland fires in Indonesia?

The answers for this sub research question were summarised in Table 2.1. It is widely know that forest fire is not a new phenomena in Indonesia, especially in the Central Kalimantan province. This province has experienced repeated fires since the early 1980s. Some fire information systems have been implemented to provide a warning of the possibility of fires such as an early warning system (Ceccato et al., 2010), fire danger ratings (De Groot et al., 2007), and hotspot monitoring (Roswintiarti et al., 2016).

Most these approaches rely only on climatic conditions such as rainfall and temperature, as the parameter in predicting the forest fires. Even though the fire danger rating system (Hoffman et al., 1995; De Groot et al., 2007) expanded variables to include the condition of vegetation fuels, the unique characteristic of peatland fires still has not been captured. As mentioned in Section 2.1.1, the nature of anthropogenic fires is the most influential factor in peatland fire occurrence (Page and Hooijer, 2016). Human activities and their involvement in creating fires should be included as one of the variables in understanding and preventing peatland fires.

RQ.1.2: What method/approach might best be used to model the escalation of peatland fires in Indonesia?

This question is answered in great detail in Chapter 2. An extensive explanation of the nature of peatland fires is presented in Section 2.1. Through this review, it is known that peatland fires are not only influenced by the climatic conditions but also human involvement and the characteristics of fuel. Due to this complexity, it is essential to possess a comprehensive collection of knowledge before developing a prediction system for the escalation of peatland fires.

In Chapter 2, the review expanded its scope to include the forest fire prediction tools that are implemented in other regions. However, due to the different characteristics of fire, implementing the fire prediction tools from other regions is not suitable for Indonesia. Taylor and Alexander (2006) mentioned that it is important for each region or nation to consider building or developing their own fire management system that can accommodate the characteristics of the wildfire.

RQ.2: In the absence of sufficient historical ground truth fire escalation data, can peatland fire escalation be modelled in a quantitative manner?

The answer for this second research question is covered in Chapter 4, Chapter 5, and Chapter 6. A knowledge-based approach is used to address the lack of appropriate models for understanding the complex behaviour of peatland fires and predicting escalation of hotspots to peatland fires. This research question has two sub-research questions:

RQ.2.1: How can information from the literature be extracted to identify contributing factors for peatland fire escalation?

This sub-research question was answered through the explanation in Chapter 4. The automated process of identifying the causal variables of hotspot escalation into peatland fire is also presented as the first stage of the workflow for developing the causal model (Figure 3.2). Section 4.3 showed how the LDA algorithm (Blei and Lafferty, 2009) of topic modeling was implemented to extract topics and terms from the documents. The interpretation of the terms and topics as the variables influencing the hotspot escalation is presented in the Section 4.4. Based on the interpretation, 15 terms were extracted and classified into three categories. These categories reflect the unique characteristics of peatland fire occurrences in Indonesia: climate conditions, human activities, and biophysical conditions, as presented in Table 4.5.

The evaluation of terms resulting from this automated process was conducted using a comparison technique. Section 4.5 shows how a list of contributing factors provided by the experts was compared with the terms from the documents. The result from the comparison shows that the terms extracted from the literature using the topic modeling method are quite similar to the terms provided by the experts. This result indicates that in the situation where no domain experts are available to provide the causal variables for a causal model, automation of information extraction from the literature may be used to replace the domain experts' contribution.

RQ.2.2: How can expert knowledge be incorporated to develop a comprehensive understanding of the characteristics of peatland fires and used to predict the escalation of peatland fires?

Given that there is insufficient historical data to build a causal model, the use of expert knowledge as a knowledge source is explored in Chapter 5 and 6. The knowledge elicitation is part of Stage 2 and Stage 3 in the workflow of causal model development. The knowledge elicited from the experts is quantified and presented in a graphical model using a Bayesian Network (BN) in the development of a peatland fire model.

As presented in Chapter 5, a focus group discusion was conducted to create the structure of the causal model. The terms resulting from the first stage was used as the intial talking point of the disscussion. The initial structure was created and presented in Section 5.2. The creation process was started by defining the objective of the causal model, identifying the causal variables, and determining the relationship for each variable. The causal model is represented using a graphical model using BN, the evaluation of the structure of the causal model follow some of the BN guidelines(Korb and Nicholson, 2011; Pollino and Henderson, 2010). During the process of expert knowledge elicitation using focus group discussion, few challenges were identified and described in Section 5.3. These challenges have impacted the result of the process. However, despite all of these challenges, a comprehensive causal model for explaining the surface peatland fire escalation was established.

The variables/nodes and their relationship developed in the second stage were quantified. Chapter 6 describes how the experts' opinions were elicited to gather the probability distribution of each node. A method from the Das (2004) was implemented to populate the CPTs. Using this method, the number of questions of the probability distribution was reduced and it eased the expert's burden in answering numerous questions. An online survey was generated using the platform SurveyMonkey. Experts were expected to fill the survey. After analysis the results from the online survey, a few issues were identified and discussed in Section 6.5.

RQ.3: In the absence of sufficient historical ground truth fire escalation data and a *gold standard* model, how can the causal model be evaluated?

This research question is answered in detail in Chapter 7 and aims to describe the evaluation process in the workflow for developing the causal model 3.2. The process of evaluating the performance of the causal model is to ensure that the model is able to capture the complexity of peatland fires in Indonesia.

The evaluation was conducted over the limited amount of historical data from the KFCP area. Section 7.2 describes how the data for this valuation was prepared. The performance evaluation was conducted based on the error rate and accuracy of the causal model in prediting the escalation of hotspots into peatland fire. Based on the error rate and accuracy result, the causal model incorrectly predicted almost one-third of the dataset. However, the test result shows that the causal model delivers a good perfomance in predicting hotspot escalation into peatland fire. The threshold T=50% also was chosen to determine the probability of hotspot escalation and non-escalation hotspots. Using this threshold, the causal model delivers better predictions of hotspot escalation into peatland fire.

The sensitivity analysis was also conducted in order to ensure the causal model captured the characteristics of peatland fires. Based on four scenarios generated for the analysis, the causal model has been able to show the correct reasoning in the climatic conditions and peat ecology. The causal model is also able to quantify the human involvement that influences the escalation of hotspots. The reasoning is mostly presented in a narrative form. However, the causal model also incorrectly predicts hotspot escalation and non-escalation. After a further analysis, it was found that *overlooked influencing factors and experts were being indecisive in providing the probability for the causal model has become the factors of this inaccurate prediction.*

Since there is no gold standard model, the performance of the causal model was compared with the result from the implementation of the LAPAN's rule, as described in Section 7.3.1. The causal model showed a slightly better performance in predicting the hotspot escalation compared to the implementation of the LAPAN's rule. When predicting non-escalation of hotspots, the implementation of the LAPAN's rule shows a slightly better result. Due to there being no clear explanation on how LAPAN detects the smoke occurrence, the RGB image processing of smoke detection conducted in this research might not be as deep or comprehensive as LAPAN's image processing. Therefore, the result from the implementation of the LAPAN's rules might improve if a comprehensive RGB image processing is conducted.

8.1.3 Research Contribution

The contributions of this thesis are presented below:

1. A repeatable workflow that quantifies the causal relationship amongst the factors contributing to escalation of hotspots into peatland fires.

   This generic workflow can be applied to solve real-life phenomena with complex and uncertain problems. There are three aspects to the contribution in this workflow:

   (a) This research has provided a deeper insight into incorporating information from literature and knowledge from experts through the repeatable workflow. The factors contributing to the escalation of peatland fires and determines the causal relationship can be identified. The workflow can also deal with the complex and uncertain problem that occurs due to insufficient historical data.

   (b) An improvement using topic modeling to the general process of the development of the causal model.

   Currently, most research uses historical data, existing models, or even human judgement to identify the variables and build the structure of the causal model. In many domain areas including peatland fires in Indonesia, insufficient historical data is a major challenge to building a purely data-driven model. Therefore, as discussed in Chapter 4, the process of identifying the causal variables can be done by using topic modeling in the automated literature analysis. The lack of hard data or even the unavailability of expert judgement in providing the causal variables for a particular complex and uncertain environmental issue can be resolved using this automated literature analysis.

(c) The development of the causal model occurs in a multi-disciplinary domain with experts from different disciplines.

A cross-disciplinary collaboration was conducted to model the hotspot escalation into peatland fire. As shown in Figure 1.1, different expertise was incorporated in conducting a knowledge-driven approach for predicting hotspot escalation into peatland fire. In the knowledge elicitation process, experts from different disciplines were involved in the process. Therefore, the causal model presented in this thesis has been able to display the complexity of peatland fire characteristics, from the anthropogenic factors, climatic conditions, and ecological perspectives.

2. The findings of this research provide insights for in the use of topic modeling from the published literature to automatically extract the influencing factors of hotspot escalation. The quality of the topic modeling result is proven to be complementary to the experts' opinions. This result indicates that in the situation where no domain experts are available to provide the causal variables for a causal model, automation of information extraction from the literature may be used to replace the domain experts' contribution.

3. The insights gained from this study may be of assistance in capturing the expert's thinking process about how hotspots escalate into peatland fires through the development of a causal model.

   The causal model has been able to bridge the gap between peatland fire science and data science. For peatland fires in Central Kalimantan, Indonesia, a massive amount of domain knowledge is mostly available in narratives and qualitative results (Applegate et al., 2002; Dennis et al., 2005; Adinugroho, 2005; Vayda, 2010; Cochrane, 2010). However, not much research has been conducted to extract and utilise this available information for the modelling of causal relationships between anthropogenic factors, climatic conditions, and ecological parts of the peatland fires. The modelling process of these environmental problems most likely depends on historical data, as mostly happens in the data science domain. This research addresses the gap between the data science and the related forest fire science. Thus, the causal model of hotspot escalation presented in this thesis is developed by capturing the experts' knowledge and incorporating a limited amount of historical data into the evaluation process.

In addition to that, the causal model also can be used to support the decision making in preventing the escalation of hotspots into peatland fires. This is the practical contribution of this thesis. The causal model of hotspot escalation into peatland fires incorporates comprehensive knowledge on why a hotspot can escalate into wildfires in many different situations. This knowledge is used to predict future peatland fire escalation. As suggested by Saharjo (2016), the fire authorities need a *system that could identify the escalation of hotspots into peatland fires and deliver accurate information about the location of a peatland fire once a hotspot is detected.* When the accurate location of potential fires is obtained, preventative actions can be taken immediately.

8.2 Future Works

This work can be viewed as the first attempt of bridging the gap of forest science and data science in predicting hotspot escalation into peatland fire. Many new areas can yet be explored. This includes some adaptation, tests. Experiments have been left for the future work due to lack of time in the process of developing the causal model.

1. Generalisation of the causal model of hotspot escalation into peatland fire.

   It could be interesting to consider the condition in other regions that have similar peatland fire characteristics as Central Kalimantan. The causal model was developed under the context of peatland fire in Central Kalimantan. The model was also analysed and evaluated using the data in the KFCP Area located in Central Kalimantan. Even though, the accuracy of the model was not extraordinary, the model has been able to capture the characteristics of peatland fire. In order to measure the generalisation of this causal model, further analysis using data from other regions and also different time frames is performed.

2. Classification of the survey questions based on the experts'

   The way of structuring the questions in the online survey could be changed to improve the validity of the experts' answers. In this research, experts were expected to answer all the questions, even though the questions did not relate to their expertise. As mentioned in Chapter 6, one expert refused to answer the question outside his/her expertise and one expert gave indecisive answers. If the questions that assigned to experts are classified and shown based on the expertise of each expert, experts might be more confident and decisive in answering the survey.

3. Aggregation of the individual expert answers

   Multiple experts were involved in the development of the causal model. Various aggregation methods are available. In this research, only majority voting, weighted mean, and analytic hierarchy process methods have been explored and implemented for different parts of the development process. There is no further analysis in this research about the best aggregation methods that should be implemented.

4. Decision making tool based on the causal model.

   One of the objectives of this research is to present the qualitative analysis of peatland fires in Indonesia in the quantitative modelling. Visualising the result of the causal model into decision making tools is out of the scope of this research. However, it could be interesting to *see* how a causal model is incorporated in a interactive format to support the decision maker in taking action to prevent forest fires.

Appendix A

Ethics

Monash University Human Research Ethics Committee

Approval Certificate

This is to certify that the project below was considered by the Monash University Human Research Ethics Committee. The Committee was satisfied that the proposal meets the requirements of the *National Statement on Ethical Conduct in Human Research* and has granted approval.

Project Number:	10887
Project Title:	Expert Knowledge Elicitation via Online Survey for the Creation of a Causal Model of Peatland Surface Fire Escalation
Chief Investigator:	Dr Grace Rumantir
Expiry Date:	31/10/2022

Terms of approval - failure to comply with the terms below is in breach of your approval and the *Australian Code for the Responsible Conduct of Research*.

1. The Chief Investigator is responsible for ensuring that permission letters are obtained, if relevant, before any data collection can occur at the specified organisation.
2. Approval is only valid whilst you hold a position at Monash University.
3. It is responsibility of the Chief Investigator to ensure that all investigators are aware of the terms of approval and to ensure the project is conducted as approved by MUHREC.
4. You should notify MUHREC immediately of any serious or unexpected adverse effects on participants or unforeseen events affecting the ethical acceptability of the project.
5. The Explanatory Statement must be on Monash letterhead and the Monash University complaints clause must include your project number.
6. Amendments to approved projects including changes to personnel must not commence without written approval from MHUREC.
7. Annual Report - continued approval of this project is dependent on the submission of an Annual Report.
8. Final Report - should be provided at the conclusion of the project. MUHREC should be notified if the project is discontinued before the expected completion date.
9. Monitoring - project may be subject to an audit or any other form of monitoring by MUHREC at any time.
10. Retention and storage of data - The Chief Investigator is responsible for the storage and retention of the original data pertaining to the project for a minimum period of five years.

Thank you for your assistance.

Professor Nip Thomson

Chair, MUHREC

CC: Ms Ariesta Lestari

List of approved documents:

Document Type	File Name	Date	Version
Supporting Documentation	The sample clips from video recording	01/09/2017	V.1
Questionnaires / Surveys	1st Online Survey - Eliciting relative weights of each parent	26/10/2017	v1
Questionnaires / Surveys	2nd Online Survey-Eliciting Compatible Parent Configurations	26/10/2017	v1
Questionnaires / Surveys	Third Online Survery-Elictation Expert Survey	26/10/2017	v1
Focus Group questions	Sample of question for focus group discussion	26/10/2017	v1
Explanatory Statement	Explanatory statement	27/10/2017	v2
Consent Form	consent-form	27/10/2017	v1
Consent Form	CONSENT FOR USING THE VIDEO DOCUMENTATION OF MEETING	27/10/2017	v1
Supporting Documentation	Sample of invitation letter for the participants to join the research project	27/10/2017	v1

Appendix B

Description of each node and state

Variables	Description	States
Population density	Number of people living in a village per square kilometre. The total number of person living in the village and were obtained from the Indonesian Statistic Agency. The population density is measured using three categories or states (low, moderate and high).	**Low**: population density less than the average density of population in Central Kalimantan **Moderate**: population density about the same with average density of population in Central Kalimantan **High**: population density above the average density of population in Central Kalimantan
Distance of hotspot from the nearest settlement	The distance of a hotspot to the nearest settlement. The location of settlement will impact the people's ability to handle the fire. In the peatland area, more fires are found in the areas that further away from the population.	**Less than 1 km**: the distance of spot fires/hotspots are less than 1km from the nearest settlement **Between 1km – 4 km**: the distance of spot fires/hotspots are between 1km and 4 km from the nearest settlement **More than 4 km**: the distance of spot fires/hotspots are more than 4km further away from the nearest settlement.

Variables	Description	States
Distance of hotspot from nearest access route	The distance of a hotspot to the nearest access route such as road, canal, river or other access routes that provides access to people for moving around. This access gives the opportunity for people to change/affect the fire environment. The changes in fire environment affects the probability of fire spreading. If there is no access then it's less likely people can affect the fire environment	**Less than 1 km**: the distance of spot fires/hotspots are less than 1km from the nearest access route **Between 1km – 4 km**: the distance of spot fires/hotspots are between 1km and 4 km from the nearest access route **More than 4 km**: the distance of spot fires/hotspots are more than 4 km from the nearest access route
Activity on the land	Behaviour of people when using fire on their activity on the land; whether their actions can lead to a wildfire or not. This also implies the culture, whether they are Dayak people who know how to make sure fire will not escalate or outsiders who do not have that knowledge. It also has relation to the prescribed burning, where local people tend to do the prescribed burning (Bambang H. Sahardjo, personal communication, 3 May 2017).	**Less likely to cause wildfire**: the people know how to handle the escalation of fire when they are using fire **More likely to cause wildfire**: the people are not familiar with how to handle the escalation of fires when they are using fire

Variables	Description	States
Livelihood	A set of activities that people undertake to meet the requirements of their livelihood. It is a root node, the default value for this node is obtained from expert's knowledge.	**Agriculture** such as crop farming **Intentional fire escalation**: Intentional use of fire in traditional hunting and fishing activity. In this activity people actively not just ignite the fire but also intentionally encourage the fire escalation (Andrew P. Vayda, personal communication, 22 February 2017, recording: 00:15:45) Timber harvesting such as collecting the medium and big bark in the forest **Agroforestry**: such as rubber plantation, rattan plantation |p2.5cm|p5cm|p6cm| such as gemor collection, wild rattan collection or natural honey collection
Economic activities	The economic activities/attraction happening in one area usually attract more people to come to that area. For example, if in one area there is mining or fishing area, there is high probability of creation and escalation of fire because more people go to that area (Bambang H. Sahardjo, personal communication, 22 February 2017, recording: min 00:57:06).	**Low activity**: less economic activity is happening **Medium activity**: medium economic activity is happening **High activity**: high economic activity is happening

Variables	Description	States
Land condition	The physical and biological status of land. The condition of land also restrict the land's productivity capacity	**Degraded**: condition of land where it has lost some degree of productivity due to human-causes (such as excessive harvesting of wood or non-timber products), or has experienced repeated fire **Non-degraded**: land that still productive
Land use - land cover	The observed biophysical cover on the earth's surface. The states in this node is the simplified version of the classification from this Indonesia National Standardization Agency	**Agriculture land**: means an area that used for agriculture, agroforestry and plantation **Mixed forest/mixed shrub**: means an area that not being used in the agriculture. This area can still be in the form of forest, shrubs or savanna **Open area**: means an area that is not covered by vegetation, can be natural or human-made **Settlement**: means an area that has permanent settlement, building or road **Water body**: means the area that covered with water. Could be natural or human made
Summation of monthly rainfall previous	The summation of 30-days rainfall before a hotspot is detected in the area	**Below 100mm**: means the summation of prior monthly rainfall calculated on daily basis is less than 100mm **Above 100mm**: means the summation of prior monthly rainfall calculated on daily basis is more than 100mm

Variables	Description	States
Summation of monthly rainfall previous 30 days	The summation of 30-days rainfall, one-month gap to the hotspot occurrence in the area	**Below 100mm**: means the summation of prior monthly rainfall calculated on daily basis is less than 100mm **Above 100mm**: means the summation of prior monthly rainfall calculated on daily basis is more than 100mm
Relative humidity	Vhe amount of moisture in air	**low**: 0% - 50% **medium**:50% - 70% **high**: 70% - 100%
Local wind speed	The movement of air in a particular area	**low**: below 6 km/hour **high**: above 6 km/hour
Ground water level	The distance of groundwater from the surface (Wösten et al., 2008)	**> greater than 40cm from the surface**: means the water level is far away from the surface and the surface is drier **< 40mm from the surface**: means the water level is closer to the surface and it is wetter/moister
Peat decomposition level	The maturity level of the peat	**Safric**: peat that is fully decomposed, which less than one sixth is recognizable as original plant material **Hemic**: peat that partially decomposed between 33% - 66% decomposed **Fibric**: peat that is less than 33% decomposed, usually it still has log on it
Fire susceptibility	How likely fire is to escalate under the influence of the nearest settlement and road access	**Low** : low risk of fire susceptibility **Medium**: medium risk of fire susceptibility **High**: high risk of fire susceptibility

Variables	Description	States
Fuel flammability	The degree of flammability of each land use - land cover. This node also refers to the fuel on the surface of peatland such as vegetation, logs, braches, leaves, small piece of wood	**Low**: the fuel and its condition in the selected area reduces fire occurrence and the escalation of fire **Medium**: the fuel and its condition on the selected area has a moderate influence on fire occurrence and the escalation of fire **High**: the fuel and its condition in the selected area enhances fire occurrence and escalation of fire
Peat flammability	The level of flammability of the peat. This condition depends on the condition of ground water level. If the water level is far away from the surface, it means the peat becomes dry and flammable. However, it also depends on the peat decomposition because there are differences on the silica and ash content in the peat	Low Medium High
The degree of fire danger	The degree of difficulty of suppressing a fire once it's started	Low Moderate High
Surface peatland fire escalation	This variable represent whether there are fires detected beyond the initial hotspot within the area of x by y km2 (within period of z number of days)	**Yes**: there is possibility of fire will escalate **No**: the fire most likely will not escalate

Appendix C

Questionnaire

C.1 Questionnaire for Eliciting Compatible Parent Configuration

Eliciting Compatible Parent Configurations

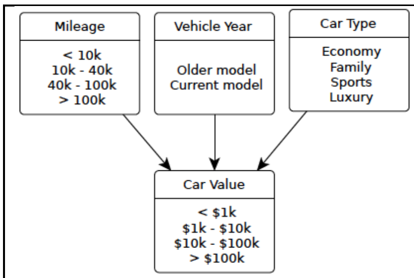
Given a parents/child or causes/effect sub model complete with the states of each variables, a domain expert will be asked to do the following:

For each state of each parent, which state of each of the other parent is ***the most likely (compatible)?***

Example:

For each of the state on the first column, please choose the most compatible state of each of the variable on the right.

Note: If you have more than 1 choice, then please list them in order of your preference. Please do this only when you feel it is absolutely necessary as it will add to the number of probabilities to elicit later.



Mileage	V. Year	Car Type
<10k	Current Model Current Model	Luxury Family
10k-40k	Current Model	Family
40k-100k	Current Model Older Model	Economy Family
>100k	Older Model	Economy

This doesn't need to elicited as all of the states CPCs can be taken from the previous answers:

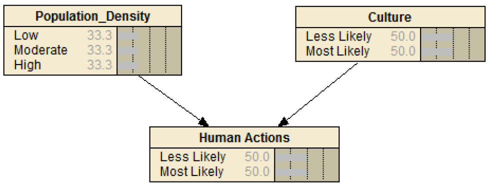
Vehicle Year	Mileage	Car Type
Older Model	>100k 40k-100k	Economy Family
Current Model	<10k <10k 10k-40k 40k-100k	Luxury Family Family Economy

Only the CPC of states Sports and Luxury of variable Car Type need to be elicited:

Car Type	Vehicle Year	Mileage
Economy	Older Model Current Model	40k-100k >100k
Family	Current Model Current Model	10k-40k <10k
Sports		
Luxury		

Sub-model Human Actions:

For each of the state on the first column, please choose the most compatible state of each of the variable on the right.

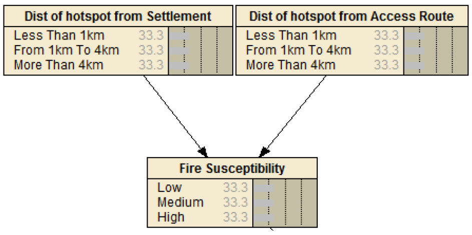


Population density	**Culture** More likely Less likely
Low density	
Moderate density	
High density	

Culture	**Livelihood** Low density Moderate density High density
More likely	
Less likely	

Sub-model Fire Susceptibility:

For each of the state on the first column, please choose the most compatible state of each of the variable on the right.

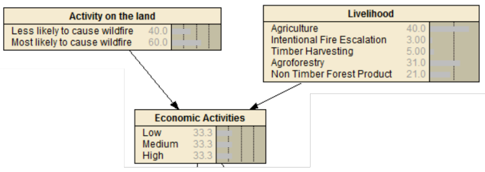


Distance of hotspot from settlement	**Distanceof hotspot from access route** Less than 1km From 1 to 4 km More than 4 km
Less than 1 km	
From 1 to 4 km	
More than 4 km	

Distance of hotspot from access route	**Distance of hotspot from settlement** Less than 1km From 1 to 4 km More than 4 km
Less than 1 km	
From 1 to 4 km	
More than 4 km	

Sub-model Economic Activities:

For each of the state on the first column, please choose the most compatible state of each of the variable on the right.

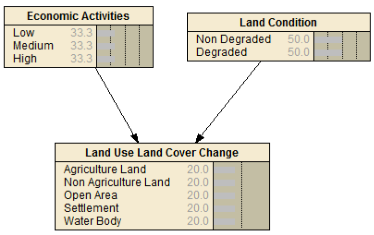


Livelihood	Activity on the land More likely Less likely
agriculture	
intentional fire	
Timber harvesting	
Agroforestry	
Non timber	

Activity on the land	Livelihood Agriculture Intentional fire Timber harvesting Agroforestry Non timber
More likely	
Less likely	

Sub-model Land use land cover:

For each of the state on the first column, please choose the most compatible state of each of the variable on the right.

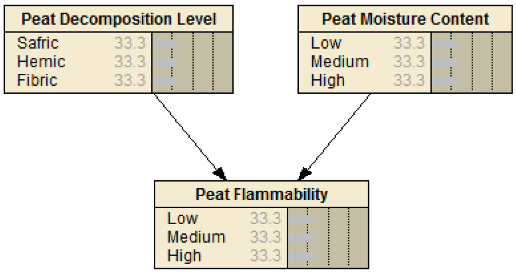


Economic activities	Land condition Non degraded degraded
Low	
Medium	
High	

Land condition	Economic activities Low Medium High
Non degraded	
degraded	

Sub-model Peat Flammability:

For each of the state on the first column, please choose the most compatible state of each of the variable on the right.

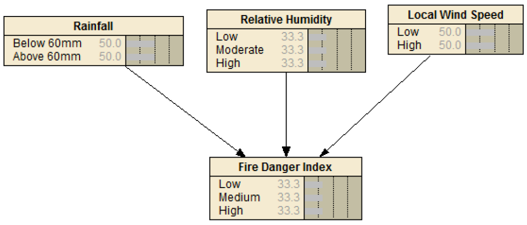


Peat decomposition	**Peat moisture content** Low Medium High
Safric	
Hemist	
Fibric	

Peat moisture content	**Peat decomposition** Safric Hemist Fibric
Low	
Medium	
High	

Sub-model The Degree of Fire Danger:

For each of the state on the first column, please choose the most compatible state of each of the variable on the right.



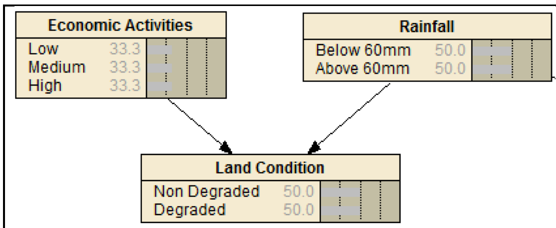
Rainfall	**Humidity** Low Medium High	**Wind_Speed** Low High
Below 60mm	high	High
Above 60mm	low	Low

Humidity	**Rainfall** Below 60mm Above 60mm	**Wind speed** Low High
Low		
Medium		
High		

Wind Speed	**Rainfall** Below 60mm Above 60mm	**Humidity** Low Medium High
Low		
High		

Sub-model Land condition

For each of the state on the first column, please choose the most compatible state of each of the variable on the right.

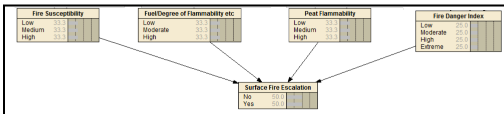


Economic activities	**Rainfall** Below 60mm Above 60mm
Low	
Medium	
High	

Rainfall	**Economic Activities** Low Medium High
Below 60mm	
Above 60mm	

Sub-model Surface Fire Escalation

For each of the state on the first column, please choose the most compatible state of each of the variable on the right.



Fire susceptibility	**Fuel/degree of flammability etc** Low fuel Moderate fuel High fuel	**Peat Flammability** Low Medium High	**Fire Danger Index** Low Medium High
Low			
Medium			
High			

Fuel/degree of flammability etc	**Fire susceptibility** Low Moderate High	**Peat Flammability** Low Medium High	**Fire Danger Index** Low Medium High
Low fuel			
Moderate fuel			
High fuel			

Peat Flammability	**Fire susceptibility** Low Moderate High	**Fuel/degree of flammability etc** Low fuel Moderate fuel High fuel	**Fire Danger Index** Low Medium High
Low			
Medium			
High			

Fire Danger Index	**Fire susceptibility** Low Moderate High	**Fuel/degree of flammability etc** Low fuel Moderate fuel High fuel	**Peat Flammability** Low Medium High
Low			
Medium			
High			

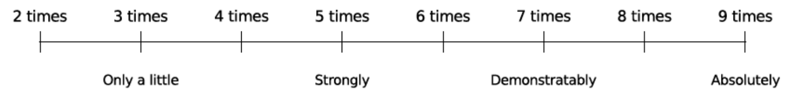
C.2 Questionnaire for Eliciting Relative Weights

Eliciting relative weights of each parent

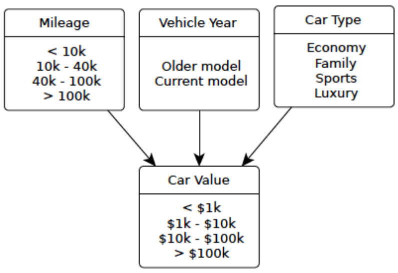
Given a parents/child or causes/effect sub model, a domain expert will be asked to do a pairwise comparison of a parent to each of the other parent and determine which one is more influential (or equally influential) to the child variable/node.

For each of the following parents/child sub model, please answer the following question:

Which parent variable has the largest influence on the child variable?
How much more influential is this parent variable against one of the other parents?



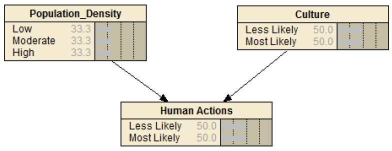
Example:



Mileage	1	V. Year	5	This means **Vehicle Year** is 5 times more influential to **Car Value** than **Mileage**
Mileage	4	C. Type	1	This means **Mileage** is 4 times more influential to **Car Value** than **Car Type**
V. Year	2	C. Type	1	

Question 1:
Which parent variable (population density or culture) has the largest influence on the human actions variable?

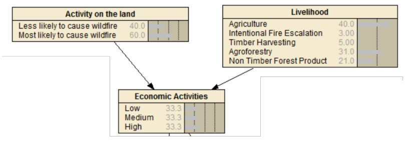
How much more influential is this parent variable against one of the other parents?



Population density		Culture	

Question 2:
Which parent variable (activity on the land, culture or livelihood) has the largest influence on the economic activities variable?

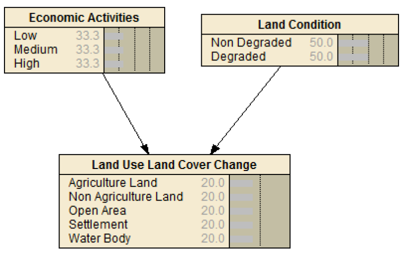
How much more influential is this parent variable against one of the other parents?



Activity on the land		Livelihood	

Question 3:
Which parent variable (economic activities or land condition) has the largest influence on the land use land cover change (LULCC) variable?

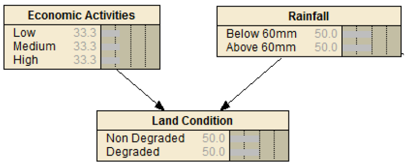
How much more influential is this parent variable against one of the other parents?



Economic Activities		Land condition	

Question 4:
Which parent variable (economic activities or rainfall) has the largest influence on the land condition variable?

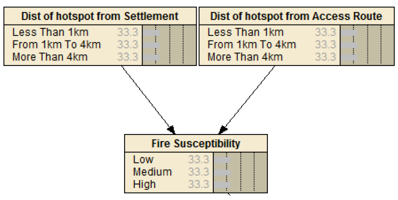
How much more influential is this parent variable against one of the other parents?



Economic Activities		Rainfall	

Question 5:
Which parent variable (distance of hotspot from settlement or distance of hotspot from access route) has the largest influence on fire susceptibility?

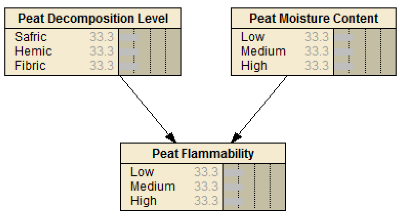
How much more influential is this parent variable against one of the other parents?



Distance of hotspot from settlement		Distance of hotspot from access route	

Question 6:
Which parent variable (peat decomposition level or peat moisture content) has the largest influence on peat flammability?

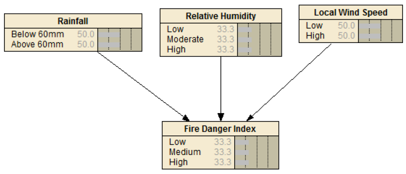
How much more influential is this parent variable against one of the other parents?



Peat decomposit ion level		Peat moisture content	

Question 7:
Which parent variable (rainfall, relative humidity, air temperature or wind speed) has the largest influence on fire danger index?

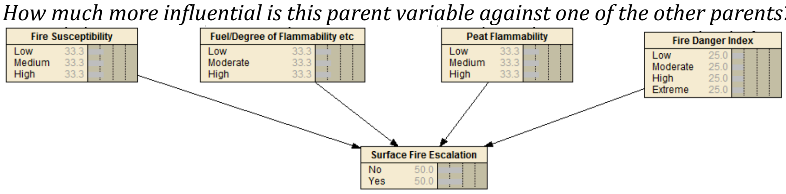
How much more influential is this parent variable against one of the other parents?



Rainfall		Relative humidity	
Rainfall		Wind speed	
Relative humidity		Wind speed	

Question 8:
Which parent variable (fire susceptibility, fuel/degree of flammability, peat flammability or fire danger index) has the largest influence on surface fire escalation?

How much more influential is this parent variable against one of the other parents?



Fire susceptibility		Fuel/degree of flammability etc.	
Fire susceptibility		Peat Flammability	
Fire susceptibility		Fire danger index	
Fuel/degree of flammability etc.		Peat Flammability	
Fuel/degree of flammability etc.		Fire danger index	
Peat Flammability		Fire danger index	

C.3 Questionnaire for Eliciting Probability Distribution

Expert Elicitation Survey

The aim of the survey is to collect information about probabilities on the relationship of influencing factors of surface peatland fire escalation in Central Kalimantan.

In the survey, you will be given a list of questions that used to elicit the conditional probabilities of each relationship in the surface peatland fire escalation model. In each questions there will be seven choices, represent the chance of a variable will be in each of its states depending on the states of its parents. Please choose the best possible chance/probability that you agree on.

Some of the questions might be out of your expertise, if so please do not skip the question and put your concerns as a comment on the “other” options box.

Rainfall Nodes

Please select the best answer to the following multiple choice questions about the probability on the influence of El Niño–Southern Oscillation (ENSO) phenomenon to the average daily amount of rainfall in Central Kalimantan:

1. What is the likelihood of the following scenario?
   "The average monthly amount of rainfall below 60mm" If we know that: El Nino is affecting the area

0% : (almost) Impossible	75%: Expected
15%: Improbable	85%: Probable
25%: Uncertain	100%: (almost) Certain
50%: Fifty-fifty	

2. What is the likelihood of the following scenario? "The average monthly amount of rainfall above 60mm" If we know that La Nina is affecting the area

0% : (almost) Impossible	75%: Expected
15%: Improbable	85%: Probable
25%: Uncertain	100%: (almost) Certain
50%: Fifty-fifty	

3. What is the likelihood of the following scenario? "The average monthly amount of rainfall above 60mm" If we know that neither El Nino or La Nina is affecting the area

0% : (almost) Impossible	75%: Expected
15%: Improbable	85%: Probable
25%: Uncertain	100%: (almost) Certain
50%: Fifty-fifty	

Relative Humidity Nodes

Please select the best answer to the following multiple choice questions about the probability on the influence of the average of rainfall's daily amount to the relative humidity.

4. What is the likelihood of the following scenario? "The relative humidity is being low" If we know that: The average monthly amount of rainfall is below 60mm

0% : (almost) Impossible	75%: Expected
15%: Improbable	85%: Probable
25%: Uncertain	100%: (almost) Certain
50%: Fifty-fifty	

5. What is the likelihood of the following scenario? "The relative humidity is being moderateIf we know that: The average monthly amount of rainfall is below 60mm

0% : (almost) Impossible	75%: Expected
15%: Improbable	85%: Probable
25%: Uncertain	100%: (almost) Certain
50%: Fifty-fifty	

6. What is the likelihood of the following scenario? "The relative humidity is being high" If we know that: The average monthly amount of rainfall is above 60mm

0% : (almost) Impossible	75%: Expected
15%: Improbable	85%: Probable
25%: Uncertain	100%: (almost) Certain
50%: Fifty-fifty	

Ground water level node
Please select the best answer to the following multiple choice questions about the probability on the influence of the daily amount of rainfall to the ground water level

7. What is the likelihood of the following scenario? "The ground water table is shallower than 40cm from the surface" If we know that: The average monthly amount of rainfall is below 60mm

0% : (almost) Impossible	75%: Expected
15%: Improbable	85%: Probable
25%: Uncertain	100%: (almost) Certain
50%: Fifty-fifty	

8. What is the likelihood of the following scenario? "The ground water table is deeper than 40cm from the surface" If we know that: The average monthly amount of rainfall is above 60mm

0% : (almost) Impossible	75%: Expected
15%: Improbable	85%: Probable
25%: Uncertain	100%: (almost) Certain
50%: Fifty-fifty	

Peat Moisture Content node
Please select the best answer to the following multiple choice questions about the probability on the influence of ground water level to the peat moisture content.

9. What is the likelihood of the following scenario? "The peat moisture content is being low" If we know that the ground water table is shallower than 40cm from the surface

0% : (almost) Impossible	75%: Expected
15%: Improbable	85%: Probable
25%: Uncertain	100%: (almost) Certain
50%: Fifty-fifty	

10. What is the likelihood of the following scenario? "The peat moisture content is being moderate" If we know that: The ground water table is deeper than 40cm from the surface

0% : (almost) Impossible	75%: Expected
15%: Improbable	85%: Probable
25%: Uncertain	100%: (almost) Certain
50%: Fifty-fifty	

11. What is the likelihood of the following scenario? "The peat moisture content is being high" If we know that: The ground water table is deeper than 40cm from the surface

0% : (almost) Impossible	75%: Expected
15%: Improbable	85%: Probable
25%: Uncertain	100%: (almost) Certain
50%: Fifty-fifty	

Peat Decomposition Level
Please select the best answer to the following multiple choice questions about the probability on the land use and land cover changes to the peat decomposition level

12. What is the likelihood of the following scenario? "The peat decomposition level is safric" If we know that: The land use and land cover over the area is agriculture land

0% : (almost) Impossible	75%: Expected
15%: Improbable	85%: Probable
25%: Uncertain	100%: (almost) Certain
50%: Fifty-fifty	

13. What is the likelihood of the following scenario? "The peat decomposition level is safric" If we know that: - The land use and land cover over the area is non-agriculture land

0% : (almost) Impossible	75%: Expected
15%: Improbable	85%: Probable
25%: Uncertain	100%: (almost) Certain
50%: Fifty-fifty	

14. What is the likelihood of the following scenario? "The peat decomposition level is safric" If we know that: The land use and land cover over the area is open areas

0% : (almost) Impossible	75%: Expected
15%: Improbable	85%: Probable
25%: Uncertain	100%: (almost) Certain
50%: Fifty-fifty	

15. What is the likelihood of the following scenario? "The peat decomposition level is safric" If we know that: The land use and land cover over the area is settlements

0% : (almost) Impossible	75%: Expected
15%: Improbable	85%: Probable
25%: Uncertain	100%: (almost) Certain
50%: Fifty-fifty	

16. What is the likelihood of the following scenario? "The peat decomposition level is safric" If we know that: The land use and land cover over the area is water bodies (eg. lakes, rivers, or canals)

0% : (almost) Impossible	75%: Expected
15%: Improbable	85%: Probable
25%: Uncertain	100%: (almost) Certain
50%: Fifty-fifty	

Settlement

Please select the best answer to the following multiple choice questions about the probability on the influence of the population density to how far a hotspot occurs from nearest settlement

17. What is the likelihood of the following scenario? "There is a settlement located less than 1 km from a hotspot" If we know that: The settlement has a low population density

0%: (almost) Impossible	75%: Expected
15%: Improbable	85%: Probable
25%: Uncertain	100%: (almost) Certain
50%: Fifty-fifty	

18. What is the likelihood of the following scenario? "There is a settlement located less than 1 km from a hotspot If we know that: The settlement has a high population density

0%: (almost) Impossible	75%: Expected
15%: Improbable	85%: Probable
25%: Uncertain	100%: (almost) Certain
50%: Fifty-fifty	

Access route

Please select the best answer to the following multiple choice questions about the probability on the influence of the population density to how far a hotspot occurs from nearest access route (eg. roads, rivers or canals)

19. What is the likelihood of the following scenario? "There is an accessible route (eg. road, river or canal) located less than 1 km from a hotspot" If we know that: The settlement has a low population density nearby

0%: (almost) Impossible	75%: Expected
15%: Improbable	85%: Probable
25%: Uncertain	100%: (almost) Certain
50%: Fifty-fifty	

20. What is the likelihood of the following scenario? "There is an accessible route (eg. road, river or canal) located less than 1 km from a hotspot" If we know that: The settlement has a high population density nearby

0%: (almost) Impossible	75%: Expected
15%: Improbable	85%: Probable
25%: Uncertain	100%: (almost) Certain
50%: Fifty-fifty	

Human Actions
Please select the best answer to the following multiple choice questions about the probability on the influence of the population density and people activity in the land to the human actions on starting fires in their activities

21. What is the the likelihood of the following scenario? "The villagers will be less likely to start a fire in their activities" If we know that:
- There is a village that has a low population density nearby
- In their culture, they are less likely to use fire.

0%: (almost) Impossible	75%: Expected
15%: Improbable	85%: Probable
25%: Uncertain	100%: (almost) Certain
50%: Fifty-fifty	

22. What is the the likelihood of the following scenario? "The villagers are being more likely to start a fire in their activities" If we know that:
- There is a village that has a low population density nearby
- In their culture, they are less likely to use fire.

0%: (almost) Impossible	75%: Expected
15%: Improbable	85%: Probable
25%: Uncertain	100%: (almost) Certain
50%: Fifty-fifty	

Economic Activities
Please select the best answer to the following multiple choice questions about the probability on the influence of the livelihood and the human actions in creating the fires to the economic activities in the area.

23. What is the likelihood of the following scenario? "The economic activity in the village is being low" If we know that:
- People's main livelihood is agriculture
- They are less likely using fire in their activities.

0%: (almost) Impossible	75%: Expected
15%: Improbable	85%: Probable
25%: Uncertain	100%: (almost) Certain
50%: Fifty-fifty	

24. What is the likelihood of the following scenario? "The economic activity in the village is being low" If we know that:
- People are intentionally using fire to do the livelihood activities
- They are more likely to use fire in their activities

0%: (almost) Impossible	75%: Expected
15%: Improbable	85%: Probable
25%: Uncertain	100%: (almost) Certain
50%: Fifty-fifty	

25. What is the likelihood of the following scenario? "The economic activity in the village is being low" If we know that:
- People's main livelihood is harvesting timber products (e.g. collecting branches, gemor

collection)
- They are more likely to use fire in their activities

0%: (almost) Impossible	75%: Expected
15%: Improbable	85%: Probable
25%: Uncertain	100%: (almost) Certain
50%: Fifty-fifty	

26. What us the likelihood of the following scenario? "The economic activity in the village is being low" If we know that:
- People's main livelihood is agroforestry
- They are less likely to use fire in their activities

0%: (almost) Impossible	75%: Expected
15%: Improbable	85%: Probable
25%: Uncertain	100%: (almost) Certain
50%: Fifty-fifty	

27. What is the likelihood of the following scenario? "The economic activity in the village is being low" If we know that:
- People's main livelihood is collecting the non-timber forest products (e.g. honey, gold)
- They are more likely using fire in their activities

0%: (almost) Impossible	75%: Expected
15%: Improbable	85%: Probable
25%: Uncertain	100%: (almost) Certain
50%: Fifty-fifty	

Land Condition

Please select the best answer to the following multiple choice questions about the probability on the influences of economic activities and the amount of daily rainfall to the land condition being degraded or non-degraded.

28. What is the likelihood of the following scenario? "The land is being a non-degraded land" If we know that:
- There is not much economic activity happening in that area
- The average daily amount of rainfall is above 60mm

0%: (almost) Impossible	75%: Expected
15%: Improbable	85%: Probable
25%: Uncertain	100%: (almost) Certain
50%: Fifty-fifty	

29. What is the likelihood of the following scenario? "The land is being a non-degraded land" If we know that:
- There is a medium economic activity happening in that area
- The average daily amount of rainfall is below 60mm

0%: (almost) Impossible	75%: Expected
15%: Improbable	85%: Probable
25%: Uncertain	100%: (almost) Certain
50%: Fifty-fifty	

30. What is the likelihood of the following scenario? "The land is being a degraded land" If we know that:
- There is not much economic activity happening in that area
- The average daily amount of rainfall is above 60mm

0%: (almost) Impossible	75%: Expected
15%: Improbable	85%: Probable
25%: Uncertain	100%: (almost) Certain
50%: Fifty-fifty	

31. What is the likelihood of the following scenario? "The land is being a degraded land" If we know that:

- There is a medium economic activity happening in that area
- The average monthly amount of rainfall is below 60mm

0%: (almost) Impossible	75%: Expected
15%: Improbable	85%: Probable
25%: Uncertain	100%: (almost) Certain
50%: Fifty-fifty	

Appendix D

Explanation of Analytic Hierarchy Process

Method to collate the expert's answers:

AHP is a multiple criteria decision-making approach and was introduced by Saaty (1990) to guide decision makers rank information based on pair-wise comparison of a number of criteria.

Generally, AHP follows three major steps (Saaty, 1990):

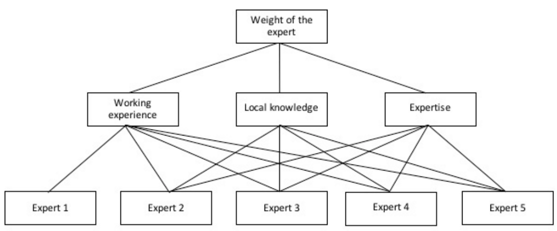
i) Establish the hierarchy of the structure. This hierarchy is used to determine the influence and impacts of the elements the goal of the problem. The goal is placed at the top level of the hierarchy; the second level or intermediate level consists of the criteria which contribute to the goals. If required, the second level can be break down into sub criteria at the next level. The lowest level of the hierarchy contains the alternatives of the decision.

ii) Elicit the pairwise comparison judgments. All the elements in the intermediate level are arranged into a matrix and elicit the judgement about the relative importance of the elements with respect to the goal.

iii) Establish the composite or global priorities of the lowest level with respect to the goal.

In this research project, the AHP method is used to calculate the weight of each expert. Four different weights are generated based on four sub-model in the causal model. The sub-models are climatic condition, human involvement, forest ecology and peatland ecology. The weight for each sub-model is used in the aggregation of conditional probability table of the nodes in the causal model.

1. *Establish the hierarchy of the structur*e. In the first or top level is the overall goal of finding the weight of the experts. In the intermediate level are the three criteria which contribute to the goal; working experience, local knowledge and expertise. The third level are the five experts which are given the weight based on the criteria in the intermediate level.



The following criteria for assessing the experts and calculating the weight for each expert:

- Working experience with forest fire, refers to how long the experts have been working or doing research in the forest fire domain. It can be determined from the year of the expert graduated of their last degree that related to their expertise.
- Local knowledge, refers to the expert's knowledge of fire in Central Kalimantan that

- Expertise, refers to skill or knowledge on a particular domain usually gained through formal training and professional practice (Martin et al., 2012). The expertise is categorised into climatic, human involvement, forest and peatland ecology.

2. Elicit the pairwise comparison judgements. In this step, all the criteria in the intermediate level are arranged into a matrix and elicit judgment of the relative importance of the criteria with respect to the overall goal. The scale the judgements based on Saaty (1990) is shown in Table 1

   Table 1 The fundamental scale

Intensity of importance scale	Definition	Explanation
1	Equal importance	Two elements or criteria contribute equally to the goal
3	Moderate importance of one over another	An element or criteria strongly favour over another based on the experience and judgment
5	Essential or strong importance	An element or criteria strongly favour over another based on the experience and judgment
7	Very strong importance	An element or criteria is strongly favoured and its dominance demonstrated in practice
9	Extreme importance	An element or criteria is extremely importance or preferred over another.
2, 4, 6, 8	Intermediate values between the two adjacent judgements	When compromised is needed

   In the pairwise comparison, the criteria are set into a pair and the importance of each pair is compared with the respect to the overall goal of generating the weight for the experts. The question to ask when comparing the criteria in the intermediate level is: of the two criteria being compared which is considered more important with the respect to the overall goal.

   In this case, there are three pairs of criteria that need to be compared. First is working experience compared to local knowledge; second is working experience compared to expertise and third is local knowledge and expertise. The Table 2 is the explanation of the comparison of each pair and the weight that generate to the criteria.

   Table 2 Explanation of the pair comparison for level 1

Working experience	1	Local Knowledge	3	The experience of working closely with forest fire provides a better understanding on the behaviour of fire. However, since the unique characteristic of peatland fire in Central Kalimantan, the local knowledge or real-life experience with particular fire should be more

				important compared to the working experience with forest fire in general. The local knowledge is judged as moderately more important than working experience. Weight: 3
Working experience	1	Expertise	5	As mention above, working experience with forest fire is important. However, forest fire is a complex problem because it is influenced by various factors. It is uncommon to find someone that experts in all aspects of fire. Being an expertise of a particular domain is essential compared to working experience. Weight: 5
Local Knowledge	1	Expertise	3	As mentioned peatland fire in Central Kalimantan is a complex problem that covers different aspect. Having experts that expertise in specific domain area will deliver a great contribution in constructing the model. Thus, it can be assumed that expertise is moderately more important compare to local knowledge. Weight: 3

The matrix of pairwise comparisons for the three criteria, along with the resulting vector of priorities is shown in Table 3. The vector of priorities is the principal eigenvector of the matrix. It gives the relative priority of the criteria measured on a ratio scale. In this case expertise has the highest priority with 63.7% influence, followed by local knowledge and working experience with 25.8% and 10.5% influence, respectively.

Table 3 Pairwise comparison matrix for level 1

	Working experience	Local knowledge	Expertise	Priority
Working experience	1	1/3	1/5	0.105
Local Knowledge	3	1	1/3	0.258
Expertise	5	3	1	0.637
Consistency Ratio (CR) = 0.04				

After the pairwise comparison matrix for the intermediate level is done, next is to generate the pairwise comparison of elements in the lowest level. The elements to be compared pairwise are the experts with respect to how much better one is than the other is satisfying each criterion in the intermediate level. Thus there will be six matrices of judgements which are working experience, local knowledge, expertise on climatic conditions, expertise on human involvement study, expertise on fuel and forest ecology and expertise on peat ecology. This is a 5 x 5 matrix since there are five experts to be pairwise compared for each other. To understand the judgement of each matrix, a brief description of the experts follows in table

Table 4 Summary of the expert's background

Expert 1	Expert 1 has worked with peatland fire for more than 20 years. This expert is a local fire expert, who lives and stays in Central Kalimantan, Indonesia. Expert 1 has a PhD on the characteristic of tropical peatland fire. The expert is now focussing the research on the community empowerment to prevent the peatland fire.
Expert 2	Expert 2 has worked with forest fire in Indonesia for more than 20 years. Even this expert is not originally from Central Kalimantan, the expert involves in different projects related to peatland fire in this area. Expert 2 completed a PhD degree in on tropical forest resource and environment. The expert is now focussing the research on the carbon emission resulted from the peatland fire.
Expert 3	Expert 3 has worked and done research about the peatland ecology and fire in Central Kalimantan since 2007. This expert is a restoration ecologist and has a PhD on investigating the barriers hindering the regeneration of degraded tropical peat swamp forest. This expert lives in Central Kalimantan now and closely involves with the local community for the past few years.
Expert 4	This expert has an expertise in climatology, geography and meteorology and also forest fire in Australia for more than 20 years. However, this expert but has limited experience and accessed to the peatland fire in Central Kalimantan.
Expert 5	Expert 5 has worked with peatland fire for 8 years. This expert is also originally from Central Kalimantan and lives in this area. Expert 5 has a PhD about the climatic condition in Indonesian and its relationship with forest fire.

Working Experience
A working experience of an expert can be determined from how long an expert has been working in a particular domain. In this research project, the working experience is related to how long has the expert worked in forest fire domain in general. The longer and expert works in forest fire domain, the higher weight will be assigned to the expert.

Table 5 Experts compared with respect to working experience

Expert 1	1	Expert 2	1	Both of the experts work in forest fire domain for more than 20 years. Weight: 1
Expert 1	5	Expert 3	1	Expert 3 has worked in forest fire domain for less than 10, however expert 1 has worked longer. Thus, Expert 1 experience is essential compare to expert 3. Weight: 5
Expert 1	1	Expert 4	1	Both of the experts work in forest fire domain for more than 20 years. Weight: 1
Expert 1	5	Expert 5	1	Expert 5 has worked in forest fire domain for less than 10, however expert 1 has worked longer. Thus, Expert 2 experience is essential compare to expert 5. Weight: 5

Expert 2	5	Expert 3	1	Expert 3 has worked in forest fire domain for less than 10, however expert 1 has worked longer. Thus, Expert 2 experience is essential compare to expert 3. Weight: 5
Expert 2	1	Expert 4	1	Both of the experts work in forest fire domain for more than 20 years. Weight: 1
Expert 2	5	Expert 5	1	Expert 5 has worked in forest fire domain for less than 10, however expert 2 has worked longer. Thus, Expert 2 experience is essential compare to expert 5. Weight: 5
Expert 3	1	Expert 4	5	Expert 3 has worked in forest fire domain for less than 10, however expert 4 has worked longer. Thus, Expert 4 experience is essential compare to expert 3. Weight: 5
Expert 3	1	Expert 5	1	Both of the experts work in forest fire domain for less than 10 years. Weight: 1
Expert 4	5	Expert 5	1	Expert 3 has worked in forest fire domain for less than 10, however expert 4 has worked longer. Thus, Expert 4 experience is essential compare to expert 3. Weight: 5

Based on the information about all of the experts, it is known that E1, E2 and E4 have work in forest fire domain longer than E3 and E4. The type, characteristic and location of fire are not part of the criteria. This criterion only focused on how long the experts have the experience and studying about forest fire in general.

The judgement in Table 5 is transferred to an AHP matrix and the priorities are calculated. Below is the calculation of each expert for the working experience criterion.

Table 6 AHP matrix of working experience criterion

	E1	E2	E3	E4	E5	Priority
E1	1	1	5	1	5	0.294
E2	1	1	5	1	5	0.294
E3	1/5	1/5	1	1/5	1	0.059
E4	1	1	5	1	5	0.294
E5	1/5	1/5	1	1/5	1	0.059
Consistency Ratio (CR) = 0						

Local knowledge

Local knowledge is the knowledge that people in a given community have developed over time. The knowledge can be obtained by the experts by the direct experience of living or interact with the community.

Table 7 Experts compared with respect to local knowledge

Expert 1	2	Expert 2	1	Expert 1 is originally from Central Kalimantan and lives in this area, while Expert 2 is not a local people. However, Expert 2 also has experience in working with the fire communities in Central Kalimantan. Thus, Expert 1 local knowledge is slightly essential compare to Expert 2 Weight: 2

Expert 1	3	Expert 3	1	As explained above, Expert 1 has a lot of experience with fire communities in Central Kalimantan. While Expert 3 is not local people and just move to this province in less than 10 years. Thus, Expert 1 experience is moderately essential compare to expert 3. Weight: 3
Expert 1	9	Expert 4	1	Expert 4 has limited experience in working with fire communities in Central Kalimantan. Expert 1 knowledge should extremely important compare to Expert 4. Weight: 9
Expert 1	3	Expert 5	1	Both of these experts are originally from Central Kalimantan. However, Expert 5 has less experience in working with the local fire communities. Thus, the Expert 1 local knowledge is moderately essential compare to expert 5. Weight: 3
Expert 2	2	Expert 3	1	Both of the expert are not originally from Central Kalimantan. Expert 3 now lives in this province, but Expert 2 has worked with the fire communities in this province longer Expert 3. Thus, Expert 2 local knowledge is slightly essential. Weight: 2
Expert 2	7	Expert 4	1	Expert 4 has limited experience in working with fire communities in Central Kalimantan. Expert 1 knowledge should strongly important compare to Expert 4. Weight: 7
Expert 2	2	Expert 5	1	Expert 5 is originally from Central Kalimantan. But Expert 2 has worked with the fire communities in this province longer than Expert 5. Thus, Expert 2 experience is slightly essential compare to expert 5. Weight: 2
Expert 3	7	Expert 4	1	Expert 4 has limited experience in working with fire communities in Central Kalimantan. Expert 3 knowledge should strongly important compare to Expert 4. Weight: 7
Expert 3	2	Expert 5	1	Both of the experts lives in Central Kalimantan provinces. However, Expert 3 has worked with the local fire communities more often compare to Expert 5. Weight: 2
Expert 4	1	Expert 5	7	Expert 4 has limited experience in working with fire communities in Central Kalimantan. Expert 5 knowledge should strongly important compare to Expert 4. Weight: 7

The judgement in table above is transferred to an AHP matrix and the priorities are calculated. Below is the calculation of each expert for the working experience criterion.

Table 8 AHP matrix of local knowledge criterion

	E1	E2	E3	E4	E5	Priority
E1	1	2	3	9	3	0.412
E2	1/2	1	2	7	2	0.251
E3	1/3	1/2	1	7	2	0.153
E4	1/9	1/7	1/7	1	1/7	0.03
E5	1/3	1/2	1/2	7	1	0.153
Consistency ratio (CR) = 0.25						

For this criteria, expert 1 is assigned with the highest weight. This is because, this expert is a local fire expert originally from Central Kalimantan and has a lot of experience working with the local communities in Central Kalimantan. Expert 5 also originally from Central Kalimantan, however this expert has less experience working the fire communities compare to Expert 2. Expert 4 is given the lowest weight in this criteria because the limited experience of fire in Central Kalimantan and also this expert is not originally from Central Kalimantan.

Expertise

Expertise of an expert refers to skill or knowledge on a particular domain usually gained through formal training and professional practice (Martin et al., 2012). The expertise is categorised into climatic, human involvement, forest and peatland ecology.

For expertise criterion, there are four weights that will be assigned to each expert. The weights for climatic condition, peat ecology, fuel flammability and human access.

Climatic Condition

The experts are then evaluated based on their expertise in the climatic condition as one of the contributing factors in forest fire. Below is the summary of the judgement of the expert's expertise in the climatic factors based on the information in Table 4.

Table 9 Experts compared with respect to expertise- climatic condition

Expert 1	1	Expert 2	2	Both experts do not have formal knowledge in climatic condition. However, Expert 2 has some publications related to climatic condition during fires in Indonesia. Thus, Expert 2 opinion slightly important compared to expert 1. Weight: 2
Expert 1	1	Expert 3	1	Both of this experts are having no expertise in climatic condition.
Expert 1	1	Expert 4	9	Expert 4 has skills and knowledge in climatic condition and has long working experience in this field. Thus, Expert 4 opinion is extremely important compared to expert 1. Weight: 9
Expert 1	1	Expert 5	7	Expert 5 has skills and knowledge in climatic condition of forest fire in Indonesia. Thus, Expert 5 opinion is strongly important compared to expert 1. Weight: 7
Expert 2	2	Expert 3	1	Both experts do not have formal knowledge in climatic condition. However, Expert 2 has some publications related to climatic condition during fires in Indonesia. Thus, Expert 2 opinion slightly important compared to expert 1. Weight: 2
Expert 2	1	Expert 4	7	Expert 2 has some publications related to climatic condition during fires in Indonesia. While Expert 4 has skills and knowledge in climatic condition and has long working experience in this field. Thus, Expert 4 opinion is strongly important compared to expert 2. Weight: 7

Expert 2	1	Expert 5	5	Expert 2 has some publications related to climatic condition during fires in Indonesia. While Expert 5 has skills and knowledge in climatic condition for Indonesian forest fire. Thus, Expert 5 opinion is important compared to expert 2. Weight: 5
Expert 3	1	Expert 4	9	Expert 4 has skills and knowledge in climatic condition and has long working experience in this field. While expert 3 has no formal knowledge on this domain. Thus, Expert 4 opinion is extremely important compared to expert 3. Weight: 9
Expert 3	1	Expert 5	7	Expert 4 has skills and knowledge in climatic condition for Indonesian forest fire. While expert 3 has no formal knowledge on this domain. Thus, Expert 4 opinion is extremely important compared to expert 3. Weight: 7
Expert 4	3	Expert 5	1	Both of the experts has formal knowledge on climatic condition during fires. But, Expert 4 has more field experience Weight: 3

The judgement in Table 9 is transferred to an AHP matrix and the priorities are calculated. As we can see inTable 10, expert 4 is assigned with the highest priority. Expert 4 background is a climatology and meteorology and has long period of time working with forest fire. This the reason of the opinion from expert 4 is extremely important compared to other experts. Expert 5 also has formal education and knowledge related to climatology in Indonesian forest fire, however compare to expert 4, expert 5 has less experiences. This the reason of opinion from expert 5 is strongly important compare to other experts, except for expert 4.

Table 10 AHP matrix of expertise-climatic condition criterion

	E1	E2	E3	E4	E5	Priority
E1	1	1/2	1	1/9	1/7	0.044
E2	2	1	2	1/9	1/5	0.072
E3	1	1/2	1	1/9	1/7	0.044
E4	9	9	9	1	3	0.557
E5	7	5	5	1/3	1	0.283
Consistency Ratio (CR) = 0.26						

Peat Ecology

The experts are then evaluated based on their expertise in the peat ecology as one of the contributing factors in forest fire. Below is the summary of the judgement of the expert's expertise in the climatic factors based on the information in Table 4

Table 11 Experts compared with respect to expertise- peat ecology

Expert 1	1	Expert 2	1	Both experts have skills and formal knowledge in peatland ecology. Weight: 1
Expert 1	1	Expert 3	1	Both experts have skills and formal knowledge in peatland ecology. Weight: 1
Expert 1	9	Expert 4	1	Expert 1 has skills and knowledge about peat ecology, while expert 4 has limited knowledge about fire in

				peatland area. Thus, Expert 1 opinion is extremely important compared to expert 4. Weight: 9
Expert 1	7	Expert 5	1	Expert 1 has skills and knowledge about peat ecology. Expert 5 has no formal knowledge in peat ecology, however this experts study about the climatic condition in peatland fire. Thus, Expert 1 opinion is strongly important compared to expert 5. Weight: 7
Expert 2	1	Expert 3	1	Both experts have skills and formal knowledge in peatland ecology. Weight: 1
Expert 2	9	Expert 4	1	Expert 2 has skills and knowledge about peat ecology, while expert 4 has limited knowledge about fire in peatland area. Thus, Expert 2 opinion is extremely important compared to expert 4. Weight: 9
Expert 2	1	Expert 5	5	Expert 2 has skills and knowledge about peat ecology. Expert 5 has no formal knowledge in peat ecology, however this experts study about the climatic condition in peatland fire. Thus, Expert 2 opinion is strongly important compared to expert 5. Weight: 7
Expert 3	1	Expert 4	9	Expert 3 has skills and knowledge about peat ecology, while expert 4 has limited knowledge about fire in peatland area. Thus, Expert 3 opinion is extremely important compared to expert 4. Weight: 9
Expert 3	1	Expert 5	7	Expert 3 has skills and knowledge about peat ecology. Expert 5 has no formal knowledge in peat ecology, however this expert study about the climatic condition in peatland fire. Thus, Expert 3 opinion is strongly important compared to expert 5. Weight: 7
Expert 4	3	Expert 5	1	Both of the experts has no formal knowledge on peatland ecology. But, expert 5 studied about the climatic condition in peatland fire. Thus, Expert 5 opinion is slightly essential compare to expert 4. Weight: 2

The judgement in Table 11 is transferred to an AHP matrix and the priorities are calculated. The matrix of pairwise comparisons for the five experts, along with the resulting vector of priorities is shown in table below. In this case, three experts shared the same priority. Expert 1, expert 2 and expert 3 have the highest priority with 30.7% influence. This means, the opinion of these experts is more important compared to expert 4 and 5. Since expert 4 and 5 has limited knowledge on peat ecology, this relative priority for these expert only 4.8% and 3.1% influence, respectively.

Table 12 AHP matrix of expertise-peat ecology criterion

	E1	E2	E3	E4	E5	Priority
E1	1	1	1	9	7	0.307
E2	1	1	1	9	7	0.307
E3	1	1	1	9	7	0.307
E4	1/7	1/7	1/7	1	0.5	0.031
E5	1/5	1/5	1/5	2	1	0.048
Consistency Ratio (CR) = 0.005						

Fuel Flammability

The experts are then evaluated based on their expertise in the forest ecology especially on the flammability of surface fuel as the judgemane one of the contributing factors in the forest fire. Below is the summary of the judgement of the expert's expertise in the climatic factors based on the information in Table 4.

The matrix of pairwise comparisons for the five experts, along with the resulting vector of priorities is shown in table below. The vector of priorities is the principal eigenvector of the matrix. It gives the relative priority of the experts measured on a ratio scale. In this case expert 2 has the highest priority with 53.5% influence; followed by expert 2 and expert 3 with 16.1% influence; expert 4 and expert 5 with, 5.5% influence and 8.8%, respectively.

	E1	E2	E3	E4	E5	Priority
E1	1	1/5	1	3	3	0.161
E2	5	1	5	5	5	0.535
E3	1	1	1	3	3	0.161
E4	1/3	1/5	1/3	1	1/3	0.055
E5	1/3	1/5	1/3	3	1	0.088
Consistency Ratio (CR) = 0.78						

Human involvement

The experts are then evaluated based on their expertise in the human involvement as one of the contributing factors in the forest fire. Below is the summary of the judgement of the expert's expertise in the climatic factors based on the information in Table 4.

The matrix of pairwise comparisons for the five experts, along with the resulting vector of priorities is shown in table below. The vector of priorities is the principal eigenvector of the matrix. It gives the relative priority of the experts measured on a ratio scale. In this case expert 1 has the highest priority with 57.6% influence, followed by expert 2 and expert 3 with 15.3% influence and expert 4 and expert 5 has the same relative priority, 5.9% influence.

	E1	E2	E3	E4	E5	Priority
E1	1	5	7	7	7	0.576
E2	1/5	1	1	5	5	0.153
E3	1/7	1	1	5	5	0.153
E4	1/7	1/5	1/5	1	1	0.059
E5	1/7	1/5	1/5	1	1	0.059

Consistency Ratio (CR) = 0.79

Synthesizing the final weight:

As mentioned above that there are four categories of expertise which are climatic, human access, peat flammability and fuel flammability. All the questions in the online survey are categorised under these category and the experts' answer for each question will be weight based on the three criteria.

The calculation of the weight is:

- The priority with the respect to working experience of each expert, multiplied by the working experience's priority and
- The priority with the respect to local knowledge of each expert, multiplied by the local knowledge's priority and
- The priority with the respect to expertise of each expert for each category, multiplied by the expertise's priority
- Sum up the total of the calculation above.

Here is the overall weight for all experts for each category:

- Climatic condition

Experts	Working experience	Local knowledge	Experise - Climate	Goal - Weight
Expert 1	0.03087	0.106296	0.028028	0.165
Expert 2	0.03087	0.064758	0.045864	0.141
Expert 3	0.006195	0.039474	0.028028	0.074
Expert 4	0.03087	0.00774	0.354809	0.393
Expert 5	0.006195	0.039474	0.180271	0.226

- Peat

Experts	Working experience	Local knowledge	Experise - Peat	Goal - Weight
Expert 1	0.03087	0.106296	0.195559	0.333
Expert 2	0.03087	0.064758	0.195559	0.291
Expert 3	0.006195	0.039474	0.195559	0.241
Expert 4	0.03087	0.00774	0.019747	0.058
Expert 5	0.006195	0.039474	0.030576	0.076

- Surface fuel flammability

Experts	Working experience	Local knowledge	Experise - Fuel	Goal - Weight
Expert 1	0.03087	0.106296	0.102557	0.240
Expert 2	0.03087	0.064758	0.340795	0.436
Expert 3	0.006195	0.039474	0.102557	0.148
Expert 4	0.03087	0.00774	0.035035	0.074
Expert 5	0.006195	0.039474	0.056056	0.102

- Human access

Experts	Working experience	Local knowledge	Experise – Human access	Goal - Weight
Expert 1	0.03087	0.106296	0.366912	0.504
Expert 2	0.03087	0.064758	0.097461	0.193
Expert 3	0.006195	0.039474	0.097461	0.143
Expert 4	0.03087	0.00774	0.037583	0.076
Expert 5	0.006195	0.039474	0.037583	0.083

References

Adinugroho, W. C. (2005). *Manual for the control of fire in peatlands and peatland forest.* Wetlands International - Indonesia Programme.

Aditama, T. Y. (2000). Impact of haze from forest fire to respiratory health: Indonesian experience. *Respirology*, 5(2):169–174.

Aiken, S. R. (2004). Runaway fires, smoke-haze pollution, and unnatural disasters in indonesia. *Geographical Review*, 94(1):55–79.

Aldrian, E. and Dwi Susanto, R. (2003). Identification of three dominant rainfall regions within indonesia and their relationship to sea surface temperature. *International Journal of Climatology*, 23(12):1435–1452.

Alkhatib, A. A. A. (2014). A review on forest fire detection techniques. *International Journal of Distributed Sensor Networks*, 10(3).

Allison, R. S., Johnston, J. M., Craig, G., and Jennings, S. (2016). Airborne optical and thermal remote sensing for wildfire detection and monitoring. *Sensors*, 16(8):1310.

Applegate, G., Smith, R., Fox, J. J., Mitchell, A., Packham, D., Tapper, N., and Baines, G. (2002). Forest fires in indonesia: impacts and solutions. In Colfer, C. J. P. and Resosudarmo, I. A. P., editors, *Which way forward : people, forests and policymaking in Indonesia*, pages 293–308. Resources for the Futuere, Washington, DC. CIFOR, Bogor, Indonesia. Institute of Southeast Asian Studies, Singapore.

Aussenac-Gilles, N., Biébow, B., and Szulman, S. (2000). Corpus analysis for conceptual modelling. In *Workshop on Ontologies and Text, 12th International Conference EKAW 2000.* Springer Verlag.

Babbie, E. R. (1990). *Survey research methods. 2nd ed.* Wadsworth.

Baddeley, M. C., Curtis, A., and Wood, R. (2004). An introduction to prior information derived from probabilistic judgements: elicitation of knowledge, cognitive bias and herding. *Geological Society, London, Special Publications*, 239(1):15–27.

Baker, S. and Mendes, E. (2010). Evaluating the weighted sum algorithm for estimating conditional probabilities in bayesian networks. In *SEKE*, volume 2010, pages 319–324.

Bank, W. (2016). The cost of fire : an economic analysis of indonesia's 2015 fire crisis. *Indonesia sustainable landscapes knowledge*, 1.

Bell, J. and Hardiman, R. J. (1989). The third role—the naturalistic knowledge engineer. In *Knowledge elicitation: principle, techniques and applications*, pages 47–85. Springer-Verlag New York, Inc.

Bennett, R. M., Jones, J., Turk, D. C., Russell, I. J., and Matallana, L. (2007). An internet survey of 2,596 people with fibromyalgia. *BMC Musculoskeletal Disorders*, 8(1):27.

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, 55(4):77–84.

Blei, D. M. and Lafferty, J. D. (2009). Topic models. In Srivastava, A. N. and Sahami, M., editors, *Text mining: classification, clustering, and applications*, volume 10, page 34. CRC Press.

Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022.

Boneh, T., Nicholson, A. E., and Sonenberg, E. (2006). Matilda: A visual tool for modeling with bayesian networks. *International Journal of Intelligent Systems*, 21(11):1127–1150.

Boynton, P. M. and Greenhalgh, T. (2004). Hands-on guide to questionnaire research: Selecting, designing, and developing your questionnaire. *BMJ: British Medical Journal*, 328(7451):1312.

Bradburn, N. M., Sudman, S., and Wansink, B. (2004). *Asking questions: the definitive guide to questionnaire design–for market research, political polls, and social and health questionnaires*. John Wiley & Sons.

Burgman, M., Carr, A., Godden, L., Gregory, R., McBride, M., Flander, L., and Maguire, L. (2011). Redefining expertise and improving ecological judgment. *Conservation Letters*, 4(2):81–87.

Burton, A., Shadbolt, N., Rugg, G., and Hedgecock, A. (1990). The efficacy of knowledge elicitation techniques: a comparison across domains and levels of expertise. *Knowledge Acquisition*, 2(2):167–178.

Butry, D. T., Mercer, E., Prestemon, J. P., Pye, J. M., and Holmes, T. P. (2001). What is the price of catastrophic wildfire? *Journal of Forestry*, 99(11):9–17.

Buxton, M., Haynes, R., Mercer, D., and Butt, A. (2011). Vulnerability to bushfire risk at melbourne's urban fringe: the failure of regulatory land use planning. *Geographical Research*, 49(1):1–12.

Byron, N. and Shepherd, G. (1998). Indonesia and the 1997-98 el niño: fire problems and long-term solutions. *Commonwealth Forestry Review*, 77:236–236.

Cain, J. (2001). Planning improvements in natural resources management. *Centre for Ecology and Hydrology, Wallingford, UK*, 124:1–123.

Cassels, J. and Johnstone, A. (1984). The effect of language on student performance on multiple choice tests in chemistry. *Journal of chemical education*, 61(7):613.

Ceccato, P. N., Jaya, I. N. S., Qian, J., Tippett, M. K., Robertson, A. W., and Someshwar, S. (2010). Early warning and response to fires in kalimantan, indonesia. *IRI Tech. Rep.*

Chandrasekharan, C. (1998). The mission on forest fire prevention and management to indonesia and malaysia (serawak). In *Bappenas/JICA/ITTO International Cross-Sectoral Forum on Forest Fire Management in Southeast Asia*, pages 7–8.

Chang, J., Gerrish, S., Wang, C., Boyd-Graber, J. L., and Blei, D. M. (2009). Reading tea leaves: How humans interpret topic models. In *Advances in neural information processing systems*, pages 288–296.

Chen, F.-W. and Liu, C.-W. (2012). Estimation of the spatial rainfall distribution using inverse distance weighting (idw) in the middle of taiwan. *Paddy and Water Environment*, 10(3):209–222.

Chokkalingam, U. and De Jong, W. (2001). Secondary forest: a working definition and typology. *The International Forestry Review*, pages 19–26.

Chokkalingam, U., Kurniawan, I., and Ruchiat, Y. (2005). Fire, livelihoods, and environmental change in the middle mahakam peatlands, east kalimantan. *Ecology and Society*, 10(1).

Chokkalingam, U. and Suyanto, S. (2004). Fire, livelihoods and environmental degradation in the wetlands of indonesia: a vicious cycle. Technical report, CIFOR, Bogor, Indonesia.

Chretien, K. C., Elnicki, D. M., Levine, D., Aiyer, M., Steinmann, A., and Willett, L. R. (2015). What are we telling our students? a national survey of clerkship directors' advice for students applying to internal medicine residency. *Journal of Graduate Medical Education*, 7(3):382–387.

Clemen, R. T. and Winkler, R. L. (1999). Combining probability distributions from experts in risk analysis. *Risk Analysis*, 19(2):187–203.

Cochrane, M. A. (2002). Spreading like wildfire: Tropical forest fires in latin america and the caribbean. Technical report, United Nations Environment Programme (UNEP).

Cochrane, M. A. (2010). *Tropical fire ecology: climate change, land use and ecosystem dynamics*. Springer Science & Business Media.

Cochrane, M. A., Alencar, A., Schulze, M. D., Souza, C. M., Nepstad, D. C., Lefebvre, P., and Davidson, E. A. (1999). Positive feedbacks in the fire dynamic of closed canopy tropical forests. *Science*, 284(5421):1832–1835.

Cohen, J. D. and Deeming, J. E. (1985). The national fire-danger rating system: basic equations. Technical report, Pasific Southwest Forest and Range Experiment Station.

Colfer, C. J. P. (2002). Ten propositions to explain kalimantan's fires. In Colfer, C. J. P. and Resosudarmo, I. A. P., editors, *Which way forward : people, forests and policymaking in Indonesia*, chapter 14, pages 309–324. Resources for the Future, Washington, DC. CIFOR, Bogor, Indonesia. Institute of Southeast Asian Studies, Singapore.

Cooke, N. J. (1994). Varieties of knowledge elicitation techniques. *International Journal of Human-Computer Studies*, 41(6):801–849.

Cordingley, E. S. (1989). Knowledge elicitation techniques for knowledge-based systems. In Diaper, D. E., editor, *Knowledge elicitation: principle, techniques and applications*, pages 87–175. Springer-Verlag New York, Inc.

Das, B. (2004). Generating conditional probabilities for bayesian networks: Easing the knowledge acquisition problem. *CoRR cs.AI/0411034*.

De Groot, W. J., Field, R. D., Brady, M. A., Roswintiarti, O., and Mohamad, M. (2007). Development of the indonesian and malaysian fire danger rating systems. *Mitigation and Adaptation Strategies for Global Change*, 12(1):165.

Deeming, J. E. (1995). Development of a fire danger rating system for east kalimantan. *IFFM short-term report*, Document No. 08.

Dennis, R. A., Mayer, J., Applegate, G., Chokkalingam, U., Colfer, C. J. P., Kurniawan, I., Lachowski, H., Maus, P., Permana, R. P., Ruchiat, Y., et al. (2005). Fire, people and pixels: linking social science and remote sensing to understand underlying causes and impacts of fires in indonesia. *Human Ecology*, 33(4):465–504.

Diaper, D. (1989). *Knowledge elicitation: principle, techniques and applications*. Springer-Verlag New York, Inc.

Dieste, O. and Juristo, N. (2011). Systematic review and aggregation of empirical studies on elicitation techniques. *IEEE Transactions on Software Engineering*, 37(2):283–304.

DiMaggio, P., Nag, M., and Blei, D. (2013). Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of us government arts funding. *Poetics*, 41(6):570–606.

Djalante, R. and Garschagen, M. (2017). A review of disaster trend and disaster risk governance in indonesia: 1900–2015. In Djalante, R., Garschagen, M., Thomalla, F., and Shaw, R., editors, *Disaster risk reduction in Indonesia: Progress, challenges and issues*, pages 21–56. Springer.

Dohong, A. and Lilia (2008). Hydrology restoration of ex mega rice project central kalimantan through canal blocking technique: lessons learned and steps forward. In Wösten, J., Rieley, J., and Page, S. E., editors, *Restoration of tropical peatlands*, pages 125–130. Alterra - Wageningen University and Research Centre, and the EU INCO – RESTORPEAT Partnership.

Druzdel, M. and Van Der Gaag, L. C. (2000). Building probabilistic networks:" where do the numbers come from?". *IEEE Transactions on knowledge and data engineering*, 12(4):481–486.

Eden, S. (1998). Environmental issues: knowledge, uncertainty and the environment. *Progress in Human Geography*, 22(3):425–432.

Eysenbach, G. and Wyatt, J. (2002). Using the internet for surveys and health research. *Journal of Medical Internet Research*, 4(2).

Fearnside, P. M. (1997). Transmigration in indonesia: lessons from its environmental and social impacts. *Environmental Management*, 21(4):553–570.

Field, R. D., Van Der Werf, G. R., Fanin, T., Fetzer, E. J., Fuller, R., Jethva, H., Levy, R., Livesey, N. J., Luo, M., Torres, O., et al. (2016). Indonesian fire activity and smoke pollution in 2015 show persistent nonlinear sensitivity to el niño-induced drought. *Proceedings of the National Academy of Sciences*, 113(33):9204–9209.

Finkel, A. M. (1996). Comparing risks thoughtfully. *Risk*, 7:325.

Flores, M. J., Nicholson, A. E., Brunskill, A., Korb, K. B., and Mascaro, S. (2011). Incorporating expert knowledge when learning bayesian network structure: A medical case study. *Artificial Intelligence in Medicine*, 53(3):181 – 204.

Forsyth, T. (2014). Public concerns about transboundary haze: A comparison of indonesia, singapore, and malaysia. *Global Environmental Change*, 25:76–86.

Frankenberg, E., McKee, D., and Thomas, D. (2005). Health consequences of forest fires in indonesia. *Demography*, 42(1):109–129.

Frey, J. H. and Fontana, A. (1991). The group interview in social research. *The Social Science Journal*, 28(2):175–187.

Friedman, N. and Goldszmidt, M. (1998). Learning bayesian networks with local structure. In *Learning in graphical models*, pages 421–459. Springer.

Fujioka, F. M., Gill, A. M., Viegas, D. X., and Wotton, B. M. (2008). Fire danger and fire behavior modeling systems in australia, europe, and north america. *Developments in Environmental Science*, 8:471–497.

Fuller, D., Jessup, T., and Salim, A. (2004). Loss of forest cover in kalimantan, indonesia, since the 1997–1998 el nino. *Conservation Biology*, 18(1):249–254.

Gable, G. G. (1994). Integrating case study and survey research methods: an example in information systems. *European Journal of Information Systems*, 3(2):112–126.

Gaveau, D. L., Salim, M. A., Hergoualc'h, K., Locatelli, B., Sloan, S., Wooster, M., Marlier, M. E., Molidena, E., Yaen, H., DeFries, R., et al. (2014). Major atmospheric emissions from peat fires in southeast asia during non-drought years: evidence from the 2013 sumatran fires. *Scientific reports*, 4.

Giglio, L., Csiszar, I., Restás, Á., Morisette, J. T., Schroeder, W., Morton, D., and Justice, C. O. (2008). Active fire detection and characterization with the advanced spaceborne thermal emission and reflection radiometer (aster). *Remote Sensing of Environment*, 112(6):3055–3063.

Giglio, L., Schroeder, W., and Justice, C. O. (2016). The collection 6 modis active fire detection algorithm and fire products. *Remote Sensing of Environment*, 178:31–41.

Goldammer, J. G. (1998). Early warning systems for the prediction of an appropriate response to wildfires and related environmental hazards. *Health Guidelines for Vegetation Fire Events, Lima, Peru*, pages 6–9.

Goldammer, J. G. and Seibert, B. (1990). The impact of droughts and forest fires on tropical lowland rain forest of east kalimantan. In *Fire in the tropical biota*, pages 11–31. Springer.

Graham, L. L. B., Masal, F., Manjin, S., Juni, E. T., Faturochman, and Applegate, G. (2014). Hotspot monitoring, fire investigation, and type and distribution of land assets in the kfcp area. Technical report, Indonesia-Australia Forest Carbon Partnership.

Greenland, S., Pearl, J., and Robins, J. M. (1999). Causal diagrams for epidemiologic research. *Epidemiology*, pages 37–48.

Griffiths, T. L. and Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National academy of Sciences*, 101(suppl 1):5228–5235.

Guo, Z., Gao, X., and Di, R. (2017). Learning bayesian network parameters with domain knowledge and insufficient data. In *Advanced Methodologies for Bayesian Networks*, pages 93–104.

Gupta, V., Lehal, G. S., et al. (2009). A survey of text mining techniques and applications. *Journal of Emerging Technologies in Web Intelligence*, 1(1):60–76.

Hardy, C. C. and Hardy, C. E. (2007). Fire danger rating in the united states of america: an evolution since 1916. *International Journal of Wildland Fire*, 16(2):217–231.

Harris, L. R. and Brown, G. T. (2010). Mixing interview and questionnaire methods: Practical problems in aligning data. *Practical Assessment, Research & Evaluation*, 15(1).

Harrison, M. E., Page, S. E., and Limin, S. H. (2009). The global impact of indonesian forest fires. *Biologist*, 56(3):156 – 163.

Hayasaka, H., Takahashi, H., Limin, S. H., Yulianti, N., and Usup, A. (2016). Peat fire occurrence. In Osaki, M. and Tsuji, N., editors, *Tropical Peatland Ecosystems*, pages 377–395. Springer.

Heil, A. and Goldammer, J. G. (2001). Smoke-haze pollution: a review of the 1997 episode in southeast asia. *Regional Environmental Change*, 2(1):24–37.

Heinrich, G. (2005). Parameter estimation for text analysis. Technical report, Technical Report.

Hennessy, K., Lucas, C., Nicholls, N., Bathols, J., Suppiah, R., and Ricketts, J. (2005). Climate change impacts on fire-weather in south-east australia. *Climate Impacts Group, CSIRO Atmospheric Research and the Australian Government Bureau of Meteorology, Aspendale.*

Herawati, H. and Santoso, H. (2011). Tropical forest susceptibility to and risk of fire under changing climate: A review of fire nature, policy and institutions in Indonesia. *Forest Policy and Economics*, 13(4):227–233.

Hickey, A. M. and Davis, A. M. (2003). Elicitation technique selection: how do experts do it? In *Requirements engineering conference, 2003. proceedings. 11th ieee international*, pages 169–178. IEEE.

Hill, T. D., Ellison, C. G., Burdette, A. M., and Musick, M. A. (2007). Religious involvement and healthy lifestyles: Evidence from the survey of texas adults. *Annals of Behavioral Medicine*, 34(2):217–222.

Hirschman, L., Hayes, W. S., and Valencia, A. (2007). Knowledge acquisition from the biomedical literature. In Baker, C. J. O. and Cheung, K.-H., editors, *Semantic Web*, pages 53–81. Springer.

Hoffman, R. R. (1987). The problem of extracting the knowledge of experts from the perspective of experimental psychology. *AI magazine*, 8(2):53.

Hoffman, R. R. and Lintern, G. (2006). Eliciting and representing the knowledge of experts. *Cambridge Handbook of Expertise and Expert Performance*, pages 203–222.

Hoffman, R. R., Shadbolt, N. R., Burton, A. M., and Klein, G. (1995). Eliciting knowledge from experts: A methodological analysis. *Organizational behavior and human decision processes*, 62(2):129–158.

Hoffmann, A. A., Schindler, L., and Goldammer, J. G. (1999). Aspects of a fire information system for east kalimantan, indonesia. In *Proceedings of the 3rd international symposium on Asian tropical forest management, Samarinda*, pages 20–23.

Hon, P. M. (1999). Singapore. In Glover, D., editor, *Indonesia's fires and haze: The cost of catastrophe*, pages 51–85. Institute of Southeast Asian Studies, Singapore.

Hoscilo, A., Page, S. E., Tansey, K. J., and Rieley, J. (2011). Effect of repeated fires on land-cover change on peatland in southern central kalimantan, indonesia, from 1973 to 2005. *International Journal of Wildland Fire*, 20(4):578–588.

Huijnen, V., Wooster, M., Kaiser, J., Gaveau, D., Flemming, J., Parrington, M., Inness, A., Murdiyarso, D., Main, B., and Van Weele, M. (2016). Fire carbon emissions over maritime southeast asia in 2015 largest since 1997. *Scientific reports*, 6:26886.

Ihsanuddin (2018). Presiden: Aturan masih sama, ada kebakaran hutan saya copot! *Kompas*.

Jensen, F. V. and Nielsen, T. D. (2007). Causal and bayesian networks. In *Bayesian Networks and Decision Graphs: February 8, 2007*, pages 23–50. Springer New York, New York, NY.

Jockers, M. L. (2014). *Text analysis with R for students of literature*. Springer.

Joosten, H. and Clarke, D. (2002). Wise use of mires and peatlands. *International Mire Conservation Group and International Peat Society*, 304.

Joosten, H., Tapio-Biström, M.-L., and Tol, S. (2012). *Peatlands: guidance for climate change mitigation through conservation, rehabilitation and sustainable use*. Food and Agriculture Organization of the United Nations.

Kinnaird, M. F. and O'Brien, T. G. (1998). Ecological effects of wildfire on lowland rainforest in sumatra. *Conservation Biology*, 12(5):954–956.

Kinseng, R. (2008). Designing system of incentive payments for environmental services in central kalimantan. *Internal Project Report. Bogor Agriculture University*.

Knol, A. B., Slottje, P., van der Sluijs, J. P., and Lebret, E. (2010). The use of expert elicitation in environmental health impact assessment: a seven step procedure. *Environmental Health*, 9(1):19.

Kontkanen, P., Myllymäki, P., Silander, T., Tirri, H., and Grunwald, P. (1997). Comparing predictive inference methods for discrete domains. In *In Proceedings of the sixth international workshop on artificial intelligence and statistics*. Citeseer.

Korb, K. B. and Nicholson, A. E. (2011). *Bayesian artificial intelligence, Second Edition*. CRC Press.

Krueger, R. A. (2014). *Focus groups: A practical guide for applied research*. Sage publications.

Kuhnert, P. M., Martin, T. G., and Griffiths, S. P. (2010). A guide to eliciting and using expert knowledge in bayesian ecological models. *Ecology letters*, 13(7):900–914.

Kunii, O. (1998). Basic fact–determining downwind exposures and their associated health effects, assessment of health effects in practice: A case study from the 1997 forest fires in indonesia. *Health Guidelines for Vegetation Fire Events, Lima, Peru*, pages 6–9.

Kunii, O., Kanagawa, S., Yajima, I., Hisamatsu, Y., Yamamura, S., Amagai, T., and Ismail, I. T. S. (2002). The 1997 haze disaster in indonesia: its air quality and health effects. *Archives of Environmental Health: An International Journal*, 57(1):16–22.

Kypri, K., Gallagher, S. J., and Cashell-Smith, M. L. (2004). An internet-based survey method for college student drinking research. *Drug and alcohol dependence*, 76(1):45–53.

Lam, L. and Suen, S. (1997). Application of majority voting to pattern recognition: an analysis of its behavior and performance. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 27(5):553–568.

Langner, A. and Siegert, F. (2006). Fires in kalimantan and sumatra 2006. *Remote Sensing Solutions GmbH*, 415.

Langner, A. and Siegert, F. (2009). Spatiotemporal fire occurrence in borneo over a period of 10 years. *Global Change Biology*, 15(1):48–62.

Laskey, K. B. and Mahoney, S. M. (2000). Network engineering for agile belief network models. *IEEE Transactions on knowledge and data engineering*, 12(4):487–498.

Leskovec, J., Rajaraman, A., and Ullman, J. D. (2014). *Mining of massive datasets*. Cambridge university press.

Lestari, A., Rumantir, G., Saharjo, B., Usup, A., Graham, L., Tapper, N., Vayda, A. P., Yulianti, N., and Teguh, R. (2018). Analysing causal factors of peatland wildfires: A knowledge-based approach. In *Proceedings of The 22nd Pasific Asia Conference on Information Systems (PACIS)*.

Li, Y., Bandar, Z. A., and McLean, D. (2003). An approach for measuring semantic similarity between words using multiple information sources. *IEEE Transactions on knowledge and data engineering*, 15(4):871–882.

Liao, W. and Ji, Q. (2009). Learning bayesian network parameters under incomplete data with domain knowledge. *Pattern Recognition*, 42(11):3046 – 3056.

Liew, S. C., Lim, O. K., Kwoh, L. K., and Lim, H. (1998). A study of the 1997 forest fires in south east asia using spot quicklook mosaics. In *Geoscience and Remote Sensing Symposium Proceedings, 1998. IGARSS'98. 1998 IEEE International*, volume 2, pages 879–881. IEEE.

Limin, S. H., Takahashi, H., Usup, A., Hayasaka, H., Kamiya, M., and Murao, N. (2007). Impacts of haze in 2002 on social activity and human health in palangka raya. *Tropics*, 16(3):275–282.

Liu, K.-R., Kuo, J.-Y., Yeh, K., Chen, C.-W., Liang, H.-H., and Sun, Y.-H. (2015). Using fuzzy logic to generate conditional probabilities in bayesian belief networks: a case study of ecological assessment. *International Journal of Environmental Science and Technology*, 12(3):871–884.

MacKinnon, K. (1996). *The ecology of Kalimantan*, volume 3. Periplus Editions.

Mahdipour, E. and Dadkhah, C. (2014). Automatic fire detection based on soft computing techniques: review from 2000 to 2010. *Artificial Intelligence Review*, 42(4):895–934.

March, S. T. and Smith, G. F. (1995). Design and natural science research on information technology. *Decision support systems*, 15(4):251–266.

Marcot, B. G., Holthausen, R. S., Raphael, M. G., Rowland, M. M., and Wisdom, M. J. (2001). Using bayesian belief networks to evaluate fish and wildlife population viability under land management alternatives from an environmental impact statement. *Forest Ecology and Management*, 153(1):29–42.

Marcot, B. G., Steventon, J. D., Sutherland, G. D., and McCann, R. K. (2006). Guidelines for developing and updating bayesian belief networks applied to ecological modeling and conservation. *Canadian Journal of Forest Research*, 36(12):3063–3074.

Margono, B. A., Potapov, P. V., Turubanova, S., Stolle, F., and Hansen, M. C. (2014). Primary forest cover loss in indonesia over 2000-2012. *Nature Climate Change*, 4(8):730–735.

Martell, D. L. (2007). Forest fire management. *Handbook of operations research in natural resources*, pages 489–509.

Martínez, J., Vega-Garcia, C., and Chuvieco, E. (2009). Human-caused wildfire risk rating for prevention planning in spain. *Journal of environmental management*, 90(2):1241–1252.

Mateus, P. and Fernandes, P. M. (2014). Forest fires in portugal: dynamics, causes and policies. In *Forest Context and Policies in Portugal*, pages 97–115. Springer.

McArthur, A. G. (1966). *Weather and grassland fire behaviour*. Forestry and Timber Bureau, Department of national Development, Commonwealth of Australia.

McArthur, A. G. (1967). *Fire behaviour in eucalypt forests*. Forestry and Timber Bureau, Department of national Development, Commonwealth of Australia.

McBride, M. F. and Burgman, M. A. (2012). What is expert knowledge, how is such knowledge gathered, and how do we use it to address questions in landscape ecology? In Perera, A. H., Drew, C. A., and Johnson, C. J., editors, *Expert knowledge and its Application in Landscape Ecology*, pages 11–38. Springer.

McCallum, A., Corrada-Emmanuel, A., and Wang, X. (2005). Topic and role discovery in social networks. In *Proceedings of 19th International Joint Conference on Artificial Intelligence*, pages 786–791. Citeseer.

McCann, R. K., Marcot, B. G., and Ellis, R. (2006). Bayesian belief networks: applications in ecology and natural resource management. *Canadian Journal of Forest Research*, 36(12):3053–3062.

Mei, Q., Shen, X., and Zhai, C. (2007). Automatic labeling of multinomial topic models. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 490–499. ACM.

Mendes, E. (2014). *Practitioner's Knowledge Representation*. Springer.

Meyer, D., Hornik, K., and Feinerer, I. (2008). Text mining infrastructure in r. *Journal of Statistical Software*, 25(5):1–54.

Meyer, M. A. and Booker, J. M. (2001). *Eliciting and analyzing expert judgment: a practical guide*. SIAM.

Miettinen, J. and Liew, S. C. (2010). Status of peatland degradation and development in sumatra and kalimantan. *Ambio*, 39(5-6):394–401.

Miettinen, J., Shi, C., and Liew, S. C. (2011). Deforestation rates in insular southeast asia between 2000 and 2010. *Global Change Biology*, 17(7):2261–2270.

Mimno, D., Wallach, H. M., Talley, E., Leenders, M., and McCallum, A. (2011). Optimizing semantic coherence in topic models. In *Proceedings of the conference on empirical methods in natural language processing*, pages 262–272. Association for Computational Linguistics.

Muhamad, N. and Rieley, J. (2002). Management of tropical peatlands in indonesia: mega reclamation project in central kalimantan. In Rieley, J. and Page, S. E., editors, *Peatlands for People, Natural Resources Function and Sustainable Management, proceedings of the International Symposium on Tropical Peatlands, August 22-23, 2001, Jakarta, Indonesia*, pages 155–160. BPPT and Indonesian Peat Association.

Mutch, R. W. (2007). *FACES: the story of the victims of Southern California's 2003 fire siege*. Wildland Fire Lessons Learned Center Tucson, AZ.

Nadkarni, S. and Nah, F. F.-H. (2003). Aggregated causal maps: An approach to elicit and aggregate the knowledge of multiple experts. *Communications of the Association for Information Systems*, 12(1):25.

Nadkarni, S. and Shenoy, P. P. (2001). A bayesian network approach to making inferences in causal maps. *European Journal of Operational Research*, 128(3):479–498.

Neuman, L. W. (2002). *Social research methods: Qualitative and quantitative approaches*. {Allyn & Bacon}.

Newman, D., Lau, J. H., Grieser, K., and Baldwin, T. (2010). Automatic evaluation of topic coherence. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 100–108. Association for Computational Linguistics.

Nicholson, A., Boneh, T., Wilkin, T., Stacey, K., Sonenberg, L., and Steinle, V. (2001). A case study in knowledge discovery and elicitation in an intelligent tutoring application. In *Proceedings of the Seventeenth conference on Uncertainty in artificial intelligence*, pages 386–394. Morgan Kaufmann Publishers Inc.

Niwattanakul, S., Singthongchai, J., Naenudorn, E., and Wanapu, S. (2013). Using of jaccard coefficient for keywords similarity. In *Proceedings of the International MultiConference of Engineers and Computer Scientists*, volume 1.

Noble, I., Gill, A., and Bary, G. (1980). Mcarthur's fire-danger meters expressed as equations. *Austral Ecology*, 5(2):201–203.

North, M., Stephens, S., Collins, B., Agee, J., Aplet, G., Franklin, J., and Fulé, P. (2015). Reform forest fire management. *Science*, 349(6254):1280–1281.

Notohadiprawiro, T. (1998). Conflict between problem-solving and optimising approach to land resources development policies: the case of central kalimantan wetlands. *The Spirit of Peatlands: 30 Years of the International Peat Society*, pages 14–24.

Nursyamsi, D., Noor, M., and Maftu'ah, E. (2016). Peatland management for sustainable agriculture. In *Tropical Peatland Ecosystems*, pages 493–511. Springer.

Offermann, P., Levina, O., Schönherr, M., and Bub, U. (2009). Outline of a design science research process. In *Proceedings of the 4th International Conference on Design Science Research in Information Systems and Technology*, page 7. ACM.

O'Hagan, A., Buck, C. E., Daneshkhah, A., Eiser, J. R., Garthwaite, P. H., Jenkinson, D. J., Oakley, J. E., and Rakow, T. (2006). *Uncertain judgements: eliciting experts' probabilities*. John Wiley & Sons.

Olesen, K. G., Kjaerulff, U., Jensen, F., Jensen, F. V., Falck, B., Andreassen, S., and Andersen, S. K. (1989). A munin network for the median nerve-a case study on loops. *Applied Artificial Intelligence an International Journal*, 3(2-3):385–403.

Organization, W. H., UNAIDS, et al. (2006). *Air quality guidelines: global update 2005*. World Health Organization.

Othman, J., Sahani, M., Mahmud, M., and Ahmad, M. K. S. (2014). Transboundary smoke haze pollution in malaysia: inpatient health impacts and economic valuation. *Environmental Pollution*, 189:194–201.

Page, S. E. and Hooijer, A. (2016). In the line of fire: the peatlands of southeast asia. *Phil. Trans. R. Soc. B*, 371(1696):20150176.

Page, S. E., Hoscilo, A., Langner, A., Tansey, K., Siegert, F., Limin, S., and Rieley, J. (2009). Tropical peatland fires in southeast asia. In *Tropical fire ecology*, pages 263–287. Springer.

Page, S. E., Siegert, F., Rieley, J., Boehm, H.-D. V., Jaya, A., and Limin, S. (2002). The amount of carbon released from peat and forest fires in Indonesia during 1997. *Nature*, 420(6911):61–65.

Payne, P. R., Mendonça, E. A., Johnson, S. B., and Starren, J. B. (2007). Conceptual knowledge acquisition in biomedicine: A methodological review. *Journal of Biomedical Informatics*, 40(5):582–602.

Pearl, J. (2000). *Causality: Models, Reasoning, and Inference*. Cambridge University Press.

Peffers, K., Tuunanen, T., Rothenberger, M. A., and Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of management information systems*, 24(3):45–77.

Pollino, C. A. and Henderson, C. (2010). Bayesian networks: A guide for their application in natural resource management and policy. *Landscape Logic, Technical Report*, 14.

Pollino, C. A., Woodberry, O., Nicholson, A., Korb, K., and Hart, B. T. (2007). Parameterisation and evaluation of a bayesian network for use in an ecological risk assessment. *Environmental Modelling & Software*, 22(8):1140–1152.

Powers, D. M. (2011). Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation. *Journal of Machine Learning Technologies*, 2(1):37–63.

Purnomo, H., Shantiko, B., Sitorus, S., Gunawan, H., Achdiawan, R., Kartodihardjo, H., and Dewayani, A. A. (2017). Fire economy and actor network of forest and land fires in indonesia. *Forest Policy and Economics*, 78:21–31.

Putra, E. I., Cochrane, M. A., Vetrita, Y., Graham, L., and Saharjo, B. H. (2016). Degraded peatlands, ground water level and severe peat fire occurrences. In *15th International Peat Congress 2016*.

Putra, E. I. et al. (2011). The effect of the precipitation pattern of the dry season on peat fire occurrence in the mega rice project area, central kalimantan, indonesia. *Tropics*, 19(4):145–156.

Rea, L. M. and Parker, R. A. (2014). *Designing and conducting survey research: A comprehensive guide*. John Wiley & Sons.

Rein, G., Cleaver, N., Ashton, C., Pironi, P., and Torero, J. L. (2008). The severity of smouldering peat fires and damage to the forest soil. *Catena*, 74(3):304–309.

Renooij, S. (2001). Probability elicitation for belief networks: issues to consider. *The Knowledge Engineering Review*, 16(3):255–269.

Rieley, J., Ahmad-Shah, A., and Brady, M. (1996). The extent and nature of tropical peat swamps. *Tropical Lowland Peatlands of Southeast Asia*, pages 17–54.

Rijksen, H. D. and Meijaard, E. (1999). *Our vanishing relative*. Tropenbos.

Rijsbergen, C. J. V. (1979). *Information Retrieval*. Butterworth-Heinemann, Newton, MA, USA, 2nd edition.

Roswintiarti, O., Parwati, Widipaminto, A., Suwarsono, Zubaidah, A., Indrajat, A., and Salyasari, N. D. (2016). Panduan teknis - v.01: Informasi titik panas (hotspot) kebakaran hutan/lahan. Technical report, CIFOR, Bogor, Indonesia.

Russo, F. et al. (2011). Explaining causal modelling. or, what a causal model ought to explain. *New Essays in Logic and Philosophy of Science, College Publications, London.*

Saaty, T. L. (1990). How to make a decision: the analytic hierarchy process. *European journal of operational research*, 48(1):9–26.

Saaty, T. L. (2008). Decision making with the analytic hierarchy process. *International journal of services sciences*, 1(1):83–98.

Saharjo, B. H. (2016). *Pengendalian kebakaran hutan dan atau lahan Indonesia.* IPB Press, Bogor, Indonesia.

Schweithelm, J. and Glover, D. (1999). Causes and impacts of the fires. In Glover, D., editor, *Indonesia's fires and haze: The cost of catastrophe*, pages 1–13. Institute of Southeast Asian Studies, Singapore.

Shahwahid H.O, M. and Othman, J. (1999). Malaysia. In Glover, D., editor, *Indonesia's fires and haze: The cost of catastrophe*, pages 51–85. Institute of Southeast Asian Studies, Singapore.

Siegert, F., Boehm, H. D. V., Rieley, J. O., Page, S. E., Jauhiainen, J., Vasander, H., and Jaya, A. (2002). Peat fires in central kalimantan, indonesia: Fire impacts and carbon release. In *International Symposium on Tropical Peatlands, Jakarta (Indonesia), 22-23 Aug 2002*. BPPT.

Silvius, M. and Diemont, H. (2007). Deforestation and degradation of peatlands. *Peatlands International*, 2(2007):32–34.

Smith, J. K., Lyon, L. J., Huff, M. H., Hooper, R. G., Telfer, E. S., Schreiner, D. S., et al. (2000). Wildland fire in ecosystems. effects of fire on fauna. *General Technical Report-RMRS-GTR-42*, 1.

(SNI), S. N. I. (2010). Land cover classification, sni no. 7645: 2010. Technical report, Indonesian National Standardization Agency.

Sonnenberg, C. and vom Brocke, J. (2011). Evaluation patterns for design science research artefacts. In *European Design Science Symposium*, pages 71–83. Springer.

Sorrensen, C. (2004). Contributions of fire use study to land use/cover change frameworks: Understanding landscape change in agricultural frontiers. *Human Ecology*, 32(4):395–420.

Steenis, M. and Fogarty, L. (2001). Determining spatial factors associated with fire ignition zones. hotspot analyses for east kalimantan. *Berau Forest Management Project, Jakarta.*

Stevens, K., Kegelmeyer, P., Andrzejewski, D., and Buttler, D. (2012). Exploring topic coherence over many models and many topics. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 952–961. Association for Computational Linguistics.

Stewart, D. W. and Shamdasani, P. N. (2014). *Focus groups: Theory and practice*, volume 20. Sage publications.

Stieger, S. and Reips, U.-D. (2010). What are participants doing while filling in an online questionnaire: A paradata collection tool and an empirical study. *Computers in Human Behavior*, 26(6):1488–1495.

Stocks, B. J., Lynham, T., Lawson, B., Alexander, M., Wagner, C. V., McAlpine, R., and Dube, D. (1989). Canadian forest fire danger rating system: an overview. *The Forestry Chronicle*, 65(4):258–265.

Stockwell, C. E., Jayarathne, T., Cochrane, M. A., Ryan, K. C., Putra, E. I., Saharjo, B. H., Nurhayati, A. D., Albar, I., Blake, D. R., Simpson, I. J., et al. (2016). Field measurements of trace gases and aerosols emitted by peat fires in central kalimantan, indonesia, during the 2015 el niño. *Atmospheric Chemistry and Physics*, 16(18):11711–11732.

Sumarga, E. (2017). Spatial indicators for human activities may explain the 2015 fire hotspot distribution in central kalimantan indonesia. *Tropical Conservation Science*, 10:1940082917706168.

Sun, L. and Yin, Y. (2017). Discovering themes and trends in transportation research using topic modeling. *Transportation Research Part C: Emerging Technologies*, 77:49–66.

Susilo, G. E., Yamamoto, K., Imai, T., Ishii, Y., Fukami, H., and Sekine, M. (2013). The effect of enso on rainfall characteristics in the tropical peatland areas of central kalimantan, indonesia. *Hydrological Sciences Journal*, 58(3):539–548.

Suwarsono, I. P. and Vetrita, Y. (2014). Zonasi daerah rawan kebakaran hutan/lahan. *Bunga Rampai Pemanfaatan Data penginderaan Jauh untuk Mitigasi Bencana*, page 51.

Suyanto, S., Sardi, I., Buana, Y., and van Noordwijk, M. (2009). Analysis of local livelihoods from past to present in the central kalimantan ex-mega rice project area. *World Agroforestry Centre, Bogor*.

Tacconi, L. (2003). Fires in indonesia: causes, costs and policy implications. Technical report, Pusat Pemanfaatan Penginderaan Jauh, LAPAN, Indonesia.

Tan, A.-H. et al. (1999). Text mining: The state of the art and the challenges. In *Proceedings of the PAKDD 1999 Workshop on Knowledge Disocovery from Advanced Databases*, volume 8, pages 65–70.

Tang, Y. Y., De Yan, C., and Suen, C. Y. (1994). Document processing for automatic knowledge acquisition. *IEEE Transactions on Knowledge and Data Engineering*, 6(1):3–21.

Taylor, S. W. and Alexander, M. E. (2006). Science, technology, and human factors in fire danger rating: the canadian experience. *International Journal of Wildland Fire*, 15(1):121–135.

Thompson, M. P., Haas, J. R., Finney, M. A., Calkin, D. E., Hand, M. S., Browne, M. J., Halek, M., Short, K. C., and Grenfell, I. C. (2015). Development and application of a probabilistic method for wildfire suppression cost modeling. *Forest Policy and Economics*, 50:249–258.

Tian, X.-r., Mcrae, D. J., Boychuk, D., Jin, J.-z., Shu, L.-f., Wang, M.-y., et al. (2005). Comparisons and assessment of forest fire danger systems. *Forestry Studies in China*, 7(1):53–61.

Tishkov, A. A. (2010). Sub-surface peat fires. *Natural Disasters-Volume II*, page 159.

Turetsky, M. R., Benscoter, B., Page, S. E., Rein, G., Van Der Werf, G. R., and Watts, A. (2015). Global vulnerability of peatlands to fire and carbon loss. *Nature Geoscience*, 8(1):11–14.

Usup, A., Hashimoto, Y., Takahashi, H., and Hayasaka, H. (2004). Combustion and thermal characteristics of peat fire in tropical peatland in central kalimantan, indonesia. *Tropics*, 14(1):1–19.

Uusitalo, L. (2007). Advantages and challenges of bayesian networks in environmental modelling. *Ecological Modelling*, 203(3):312–318.

van Beukering, P. J., Schaafsma, M., Davies, O., Oskolokaite, I., et al. (2008). The economic value of peatland resources within the central kalimantan peatland project in indonesia: perceptions of local communities.

van der Gaag, L. C., Renooij, S., Witteman, C. L., Aleman, B. M., and Taal, B. G. (1999). How to elicit many probabilities. In *Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence*, pages 647–654. Morgan Kaufmann Publishers Inc.

Van Wagner, C. (1975). Comparison of the canadian and american forest fire danger rating systems (1974).

Varis, O. (1997). Bayesian decision analysis for environmental and resource management. *Environmental Modelling & Software*, 12(2-3):177–185.

Varma, A. (2003). The economics of slash and burn: a case study of the 1997–1998 indonesian forest fires. *Ecological Economics*, 46(1):159–171.

Vayda, A. P. (1999). *Finding causes of the 1997-98 Indonesian forest fires: problems and possibilities*. WWF Indonesia.

Vayda, A. P. (2010). Explaining indonesian forest fires: Both ends of the firestick. In Bates, D. G. and Tucker, J., editors, *Human Ecology*, pages 17–35. Springer.

Vayda, A. P. (2011). Dos and don'ts in interdisciplinary research on causes of fires in tropical moist forests. In Vayda, A. P. and Walters, B. B., editors, *Causal explanations for social scientists: A reader*, pages 287–304. Lanham, MD: AltaMira Press.

Verma, T. and Pearl, J. (1991). *Equivalence and synthesis of causal models*. UCLA, Computer Science Department.

Vernimmen, R., Hooijer, A., Aldrian, E., Van Dijk, A., et al. (2012). Evaluation and bias correction of satellite rainfall data for drought monitoring in indonesia. *Hydrology and Earth System Sciences*, 16(1):133.

Vetrita, Y. and Haryani, N. S. (2011). Validasi hotspot modis indofire di provinsi riau. *Jurnal Ilmiah Geomatika*, 18(1).

Von Alan, R. H., March, S. T., Park, J., and Ram, S. (2004). Design science in information systems research. *MIS quarterly*, 28(1):75–105.

Wallach, H. M. (2006). Topic modeling: beyond bag-of-words. In *Proceedings of the 23rd International conference on Machine learning*, pages 977–984. ACM.

Wallach, H. M., Murray, I., Salakhutdinov, R., and Mimno, D. (2009). Evaluation methods for topic models. In *Proceedings of the 26th annual international conference on machine learning*, pages 1105–1112. ACM.

Whittaker, J., Haynes, K., Handmer, J., and McLennan, J. (2013). Community safety during the 2009 australian 'black saturday'bushfires: an analysis of household preparedness and response. *International Journal of Wildland Fire*, 22(6):841–849.

Wibowo, A., Suharti, M., Sagala, A., Hibani, H., and Van Noordwijk, M. (1996). Fire management on imperata grasslands as part of agroforestry development in indonesia. *Agroforestry Systems*, 36(1-3):203–217.

Wieringa, R. (2009). Design science as nested problem solving. In *Proceedings of the 4th international conference on design science research in information systems and technology*, page 8. ACM.

Winstock, A., Lynskey, M., Borschmann, R., and Waldron, J. (2015). Risk of emergency medical treatment following consumption of cannabis or synthetic cannabinoids in a large global sample. *Journal of Psychopharmacology*, 29(6):698–703.

Wooster, M., Perry, G., and Zoumas, A. (2012). Fire, drought and el niño relationships on borneo (southeast asia) in the pre-modis era (1980–2000). *Biogeosciences*, 9(1):317–340.

Wösten, J., Clymans, E., Page, S. E., Rieley, J., and Limin, S. (2008). Peat–water interrelationships in a tropical peatland ecosystem in southeast asia. *Catena*, 73(2):212–224.

Wu, S.-T., Li, Y., and Xu, Y. (2006). Deploying approaches for pattern refinement in text mining. In *Data Mining, 2006. ICDM'06. Sixth International Conference on*, pages 1157–1161. IEEE.

Wu, Z., Lei, L., Li, G., Huang, H., Zheng, C., Chen, E., and Xu, G. (2017). A topic modeling based approach to novel document automatic summarization. *Expert Systems with Applications*, 84:12–23.

Xu, J. and Croft, W. B. (1998). Corpus-based stemming using cooccurrence of word variants. *ACM Transactions on Information Systems (TOIS)*, 16(1):61–81.

Yang, X., Xie, X., Liu, D. L., Ji, F., and Wang, L. (2015). Spatial interpolation of daily rainfall data for local climate impact assessment over greater sydney region. *Advances in Meteorology*, 2015.

Yulianti, N. and Hayasaka, H. (2013). Recent active fires under el niño conditions in kalimantan, indonesia. *American Journal of Plant Sciences*, 4(3A):685–696.

Yulianti, N., Hayasaka, H., and Usup, A. (2012). Recent forest and peat fire trends in indonesia the latest decade by modis hotspot data. *Global environmental research*, 16(1):105–116.

Zagorecki, A., Voortman, M., and Druzdzel, M. J. (2006). Decomposing local probability distributions in bayesian networks for improved inference and parameter learning. In *FLAIRS Conference*, pages 860–865.

Zhang, Y. and Patrick, J. (2005). Paraphrase identification by text canonicalization. In *Proceedings of the Australasian Language Technology Workshop 2005*, pages 160–166.

Zhao, W., Chen, J. J., Perkins, R., Liu, Z., Ge, W., Ding, Y., and Zou, W. (2015). A heuristic approach to determine an appropriate number of topics in topic modeling. In *BMC bioinformatics*, volume 16, page S8. BioMed Central.