Fire Safety Journal 90 (2017) 169–180

Contents lists available at ScienceDirect

Fire Safety Journal

journal homepage: www.elsevier.com/locate/firesaf

A statistical analysis of occurrence and association between structural fire hazards in heritage housing

Yaping He*, Laurence A.F. Park

School of Computing, Engineering and Mathematics, Western Sydney University, Locked Bag 1797, Penrith 2751, Australia

ARTICLE INFO

Keywords:
Building code
Heritage
Logistic regression
Probability
Renovation
Risk
Structure

ABSTRACT

Reported in this paper is a novel application of statistical analysis of structural fire hazards that were found in heritage housing stock in a metropolitan area. The structural fire hazards in terms of non-compliances to the building regulations are digitised and then subjected to statistical analysis to obtain estimates of probabilities of occurrence under various conditions. The concepts of Hamming distance, Jaccard distance, virtual distance and pairwise Phi correlation coefficients are employed in the analysis to estimate the association between the fire hazards. Estimates of the probability distribution over the number of joint occurrence of hazards and pairwise joint probabilities are also obtained. In addition the 3-tuple and 4-tuple joint probabilities are analysed. Finally, logistic regression models are established to correlate each fire hazard with the others. The results show that not only the probability of occurrence of structural fire hazards is high, but probability of multiple occurrence is also significant. There are indications that some structural fire hazards are correlated. The findings of this study may assist certifying authorities, building surveyors, fire safety engineers and fire services in identifying fire hazards in heritage buildings and developing alleviating and effective strategies or solutions to protect life safety of building occupants as well as the cultural heritage values of the relevant building stock.

1. Introduction

Fire is one of the most frequent and common threats to public safety and social development among various kinds of disasters [1,2]. Particularly, the destruction by fire is a major threat to the conservation of heritage buildings as well as their contents worldwide [3–5]. Building fires are the result of human occupation and activity. Interestingly, it is the continuing usage and maintenance that is regarded as an effective means of conservation of heritage buildings [6,7]. Old buildings may undergo renovations or refurbishments to adapt the changes in lifestyle and technology. For example, air conditioning units or skylights may be added to existing buildings to provide comfort for building occupants or to improve energy efficiency. These kinds of renovations inevitably alter or have impacts on the structure of the existing buildings and their fire safety measures.

A distinctive feature of heritage housing in suburban areas of major metropolitan cities in Australia is the adjoined and/or close proximity of multiple properties [8]. Fire spread between these kinds of properties is a potential hazard, which needs to be addressed in order to ensure both the continued viability of the remaining heritage housing stock and the life safety of the occupants. As heritage housing is a valuable cultural and dwindling asset in most parts of the world, the level of protection against a major fire in closely-spaced heritage housing precincts deserves careful consideration.

A recent study by Hardie et al. [9] investigated the presence of a set of identified fire hazards within a sample of 47 heritage social housing properties in Sydney. The study's major contribution was the collection of data on the structural fire hazards, but unfortunately, the study did not explore deeper relationships between the hazards.

In this article, a thorough assessment of the probability of each structural fire hazard being present or absent, and the association between each hazard is provided. To complete the analysis, sophisticated statistical methods for binary random variables are used, such as confidence intervals for proportions, binary metrics, multi-dimensional scaling, and logistic regression.

This article provides the two major contributions: 1) the computation of associations between various structural fire hazards, and 2) the revelation of a given set of sophisticated analysis methods that were developed in other fields of study, that can be applied to fire hazard analysis. The article proceeds as follows: Section 2 contains a description of the problem and related work, Section 3 contains a description of the data. In Section 4, the individual fire hazards are examined. Sections 5 and 6 describe the metrics and pairwise relationships of fire hazards. Finally, Section 7 examines the effect of the set of fire hazards

* Corresponding author.
E-mail address: y.he@westernsydney.edu.au (Y. He).

http://dx.doi.org/10.1016/j.firesaf.2017.04.021
Received 5 August 2016; Received in revised form 17 March 2017; Accepted 12 April 2017
0379-7112/

Nomenclature

AIC	Akaike information criterion
C	correlation matrix
C_k	cumulative probability of joint occurrences of k multiple fire hazards
D	distance function
D_k	probability of joint occurrences of k multiple fire hazards
g	logistic link function
HD	Hamming distance
i,j,k,l	indices
JD	Jaccard distance
L	likelihood
M	total number of fire hazards
m	number of terms included in the logistic model
N	sample size or length of string
P	probability
P_j	probability of occurrence of jth fire hazard in a building
R	fire hazards
$R_{i,k}$	the realisation of hazard R_i in kth sampled building
$\bar{R}_i$	the mean of R_i over all sampled buildings
S_{Rj}	standard deviation in R_j
VD	virtual distance as defined in Eq. (5).
VS	virtual similarity
X, Y, Z	arbitrary binary strings

Greek

β_j	regression coefficient (j=0, 1, ..., M)
$\phi_{i,j}$	Phi correlation coefficient between fire hazard pair R_i and R_j
π_i	estimated probability P_i, or the marginal probability, of fire hazard i being present

on each individual fire hazard.

2. Background

Fire hazards usually appear in different forms. In the study by Hardie et al. [9], structural fire hazards are defined as the features in the building structure that are not compliant with the current building regulations, and defects in building fire safety measures as a result of poor maintenance. Fire separation between adjacent dwellings is prescribed by many building regulations as one of the fundamental strategies to prevent fire spread in close spaced properties. Hardie et al. [9] conducted an observational survey to look at the occurrence of noted defects, or non-compliance with the Building Code of Australia [10], in fire separation between attached or closely spaced occupancies. The survey selected a sample group of 47 heritage listed public housing properties in densely built up areas of Sydney. The sampled group consisted of 18 detached or stepped frontage houses and 29 row houses.

Most of these properties were built largely using brick which had less potential for fire spread than closely spaced timber housing. However, fire can spread through any significant gaps in the non-combustible walls, as well as, through any combustible material that bridges the gap between the adjacent brick buildings. The properties in the survey were inspected for any potential gaps or breaches in fire isolation and potential bridging pathways for fire between the separate occupancies, which may have occurred as a result of building refurbishments or upgrading over the lifetime of the heritage buildings. In total, ten fire hazards were identified and their frequencies of occurrence were estimated. Eight of the ten identified fire hazards were structural related [9], which are listed in Table 1 together with their estimated occurrence probabilities.

The separating walls are the walls between adjoining properties. They are required to have fire ratings by the building code [10]. When the walls are breached by penetration by combustible or non-fire rated materials or gaps (R_1 to R_4), the breaches are regarded as fire hazards. Similarly, the features denoted by R_5 to R_8 do not comply with the relevant clauses the building code and, hence, regarded as fire hazards. For detailed description of the eight structural fire hazards listed in Table 1, reference is made to the article by Hardie et al. [9]. Their study also showed that multiple fire hazards might co-exist in individual given buildings. These hazards represent the structural weaknesses and potential routes for fire spread in the event of fire.

The identified fire hazards were likely the results of building refurbishment that were undertaken without proper certification by appropriate building surveyors or the authorities having jurisdictions. Many may have been the result of unsympathetic service upgrades undertaken in a series of small refurbishments over time, which did not at the time require certification or inspection. Based on the result of their study, Hardie et al. [9] recommended that authorities having jurisdictions should consider the introduction of a building surveying audit wherever a refurbishment is undertaken in a heritage housing property, regardless of the extent of the refurbishment. Such audits will enhance heritage protection as well as life safety of building occupants.

Hazard identification is the first step in risk analysis and management [11,12]. The study by Hardie et al. [9] represents such an important step. Following this step is the systematic quantification of hazards in terms of their likelihood of occurrence. Such quantification is warranted for risk assessment, policy making and developing solutions to implement the policies and rectify the identified hazards. Recent literature has seen advancement in the Artificial Neural Networks (ANN) and Building Information Modelling (BIM) approaches to rapid risk assessment for fire service operations and fire risk management [13,14]. Statistical data of structural fire hazards in existing buildings can be of assistance to the development of ANN and BIM models. The most recent work by Silva et al. [15] demonstrated the valuable contribution of the probabilistic analysis to the evaluation of the risk of building component failure in the building lifecycle assessment. It represents the increasing trend of quantitative approach to building research.

Structural fire hazards in existing buildings, particularly in heritage buildings, are inherent defects that are deemed non-compliant to current prescriptive building regulations. Various solutions can be developed to remedy these defects. A straightforward solution could be to alter the building structures and make them compliant to the prescriptive building regulations. However, this kind of solution may run into conflict with the protection of the heritage values of the buildings. Therefore, performance based design solutions or fire safety engineering solutions under the performance based building codes are

Table 1
The eight structural fire hazards and their occurrence probability.

Fire hazard	Notation	π_i
Timber penetrations through the separating wall	R_1	0.47
Box gutters penetrating the separating wall	R_2	0.15
Gaps and other penetrations in the separating walls	R_3	0.11
Separating wall stopping short of the roof	R_4	0.21
Skylights installed within 900 mm of the adjoining property	R_5	0.17
Combustible facades bridging between attached houses	R_6	0.09
Combustible separating walls between adjacent balconies of attached houses	R_7	0.11
Common eaves construction	R_8	0.24*

* This fraction was calculated over the sampled 29 row houses. Over the total sample size of 47 (including single dwellings) this value is 0.15.

often resorted to [16,17]. Various risk based design approaches and fire safety engineering solutions have been studied. For example, the use of sprinklers as a performance based design solution for structural fire resistance was investigated by He and Grubits [18]. Huang et al. [19] studied the use of a water mist system for heritage building protection. From an engineering and building economics point of view, it would be desirable that one or limited performance based design solutions could be used to address multiple structural fire hazards. Preventing fire spread between adjacent or adjoining properties is one of the objectives of firefighting in fire incidences. Structural integrity and fire resistance capability is one of the major concerns by fire services not only for the rescue of building occupants but also for the safety of fire service personnel. Adequate anticipation of possible structural defects and fire spread routes would assist firefighters in safely and effectively carrying out their duties. For these reasons, it is, therefore, necessary to investigate not only the frequency of occurrence of structural fire hazards but also the possibility or probability of the occurrence of multiple fire hazards in adjoining dwellings and the correlations between various hazards.

In this article, a thorough assessment of the probability of each fire hazard being present in a heritage house is conducted, given knowledge of other present and absent fire hazards, or no knowledge of other fire hazards in the dwelling. To complete the analysis, sophisticated statistical methods for binary random variables are used, such as confidence intervals for proportions, binary metrics, multi-dimensional scaling, and logistic regression. The findings may assist certifying authorities, building surveyors and fire safety engineers in identifying fire hazards in heritage buildings and developing alleviating solutions to most effectively address those hazards. For example when a building surveyor is inspecting a property for any noted defects, or non-compliances (i.e. the eight fire hazards listed in Table 1) to the current building regulations, it would be desirable not only to be able to anticipate prior to the inspection the likelihood of finding a fire hazard in a heritage housing, but also to be able to anticipate, once a fire hazard is detected, the likelihood of finding other defects in the same property. Furthermore, when developing fire protection policy or fire safety engineering solutions, fire protection officials or engineers may need information to contemplate an effective approach that is capable of alleviating or eliminating multiple hazards.

3. Digitisation of raw data

In order to carry out the association analysis, the original sample of size N=47 was converted from text-based data [20] into a numerical-based matrix, as shown in Table A1 of Appendix A, that registers the identified structural fire hazards vs the audited buildings. The binary data (0, 1) is used to indicate whether a fire hazard is detected in a building. The parameters R_j ($j = 1, 2, ..., 8$) correspond to fire hazards shown in Table 1. The value 1 in the matrix represents the occurrence of a fire hazard and 0 the non-occurrence. It is noted that there were 18 detached or stepped frontage houses in the sample, to which fire hazard 'common eaves construction' is not applicable. Without losing generality, the subset of detached houses is not differentiated from the remaining sampled population in the current study. In other words, 'no common eaves construction' is simply treated as a feature of the detached houses in the same way as some of the row houses. Therefore, the detached houses are assigned '0' against fire hazard R_8 in Table A1.

In this matrix a row array F_i ($i = 1, 2, ..., 47$) indicates whether fire hazards R_j ($j = 1, 2, ..., 8$) have been detected in the ith housing property. It is the status array of a given housing property. A column array, or string, indicates how often, or the frequency of a given fire hazard R_j occurred in the sample. The column array is also denoted by R_j. The length of the column string is the sample size N (= 47). The length of the row array M (= 8) is the number of fire hazards investigated.

The proportion results as presented in Table 1 represent estimates of probability of occurrence of a given fire hazard regardless of the existence of any other fire hazards. Generally, probability can be defined as the relative frequency of event, or the proportion of a subset to the set it belongs to or to another subset it belongs to [21,22]. This definition is adopted in the current study and the estimate of probability of occurrence of fire hazards is based on evaluating the ratio of the size of a prescribed subset to that of the sampled set or another encompassing subset.

4. Probability of occurrence and probability distribution of fire hazards

The probability of occurrence, denoted by P_j, refers to a given fire hazard R_j. It is estimated by the relative occurrence frequency

$$P_j \cong \pi_j = \frac{\langle R_j \rangle}{N} \tag{1}$$

where $\langle R_j \rangle$ denotes the number of non-zero elements in string R_j, N is the cardinality, or the length of the string, or the sample size. The values of the estimates for the specifically identified hazards in the study by Hardie et al. [9] have been given in Table 1. These estimates were subjected to confidence test [23] and the results of margin of error with 95% confidence are shown in Table 2.

Note that $j = 0$ means hazard free. It can be seen from Table 2 that the 95% margins of error are significant. In order to have more accurate estimate of proportion, or probability, of fire hazard occurrence, a larger sample size would be desirable.

Without losing the preciseness, all estimates of probabilities are referred to as probabilities in the following discussions.

Table A1 reveals that multiple fire hazards may occur simultaneously in a building and yet there may be occasions that a building is free of any hazard. Let D_k ($k = 0, 1, ..., M$) denote the probability of k number of fire hazards occurring simultaneously in a building, where D_0 is the hazard free probability. D_k can be estimated as the proportion of the sample that possesses k number of fire hazards, N_k:

$$D_k = \frac{N_k}{N} \tag{2}$$

Presented in Fig. 1 is the estimated probability distribution for number of simultaneous occurrence of structural fire hazards in the sampled buildings.

The cumulative probability, C_k, as defined in the following equation

$$C_k = \sum_{i=0}^{k} D_i \tag{3}$$

is the probability of finding up to k (= 0, 1, ..., M) number of fire hazards in a heritage housing property in Sydney. The complement of C_k, i.e., $(1-C_k)$ is the probability of finding more than k number of fire hazards in a heritage housing property. Both C_k and $1-C_k$ are plotted in Fig. 2.

From Figs. 1 and 2 it can be predicted that approximately 13% of the heritage housing properties in Sydney are structurally fire hazard

Table 2
Estimated probability of occurrence of hazard j and 95% margin of error.

Index j	P_j (%)
0	12.8 ± 9.5
1	46.8 ± 14.0
2	14.9 ± 10.0
3	10.6 ± 8.6
4	21.3 ± 11.5
6	17.0 ± 10.5
6	8.5 ± 7.8
7	10.6 ± 8.6
8	14.9 ± 10.0

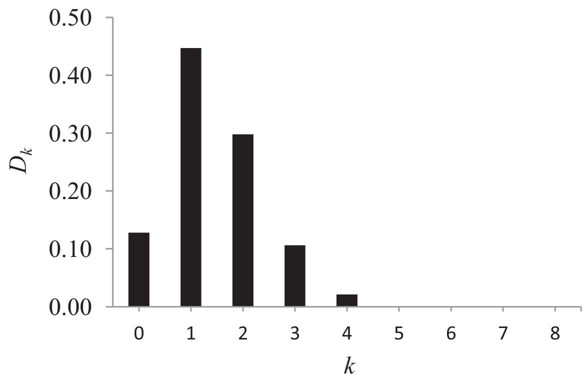

Fig. 1. Estimated probability distribution for number of simultaneous occurrence of structural fire hazards in the sampled buildings.

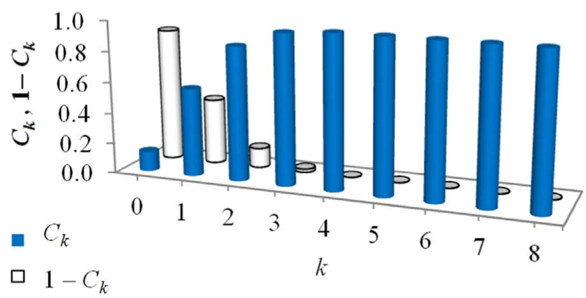

Fig. 2. Cumulative probability of multiple occurrences of fire hazards.

free. In other words, the probability of finding a heritage housing property in possession of at least one structural fire hazard is 87%. The probability of finding one fire hazard only is significantly high (~45%). The probability of simultaneous existence of 2, 3 or 4 hazards is much less and gradually decreases with the number of hazards. The probability of a heritage building possessing multiple structural fire hazards is 42% (i.e., $1-C_1=0.42$). The probability of finding more than 4 co-existing hazards in a building is approximately zero (i.e., $1-C_4=0$).

This section examined the likelihood of a fire hazard existing in a dwelling and so can assist inspectors when no other information is known. If there are known fire hazards in the dwelling, we can use the additional information to provide better estimates of fire hazard likelihood. These relationships between fire hazards are examined in the next section.

5. Probability of pairwise joint occurrence

It is possible that for a given housing property or a group of housing properties, multiple fire hazards may exist. In order to develop a strategy to alleviate the fire hazards, one may be interested in how the fire hazards are related statistically. An attempt is made in this study to employ the Hamming distance and Jaccard distance concepts to identify associations and obtain estimates of the pairwise joint probabilities between various fire hazards and of conditional probabilities. The concept of virtual distance is introduced to aid the analysis. The fundamentals of these concepts and of the evaluation methods, including the methods of Phi correlation, joint probability and logistic regression, are described in following subsections. The presentations of the results and discussions are delivered in Section 6.

5.1. Hamming distance

As descried earlier, the column arrays or strings R_j in Table A1 represent the occurrence patterns of various fire hazards. The relationship between any pair of hazards can be revealed by comparing their strings. The Hamming distance is a number used to denote the difference between two binary strings of equal length [23]. This method can be standardised by taking the difference between two binary strings of equal length and dividing the result by the length of the strings. A binary string is an ordered collection of 0's and 1's, where 1 can represent presence and 0 absence. Therefore an ordered binary string is a representation of a *set*.

Given two binary strings (or sets) of X and Y with equal length, the Hamming distance between the two strings can be expressed by the following:

$$HD(X, Y) = \frac{\langle X \cup Y \rangle - \langle X \cap Y \rangle}{N} \tag{4}$$

where for any given string Z, the brackets $\langle Z \rangle$ define the operation of counting the number of non-zero elements in the string, $X \cup Y$ and $X \cap Y$ are the union and intersection of the two strings X and Y respectively, and N is the cardinality, or the length of the strings. Further explanation of the union and intersection of binary strings is given in Appendix B.

When applied to the hazard arrays (or column arrays of structural fire hazard matrix as given in Table A1) the Hamming distance can be interpreted as a measure of the probability of finding two hazards as represented by X and Y not occurring simultaneously, or occurring mutually exclusively.

A Hamming distance $HD(X, Y)$ has the following properties:

- Symmetry: $HD(X, Y) = HD(Y, X)$,
- $HD(X, Y) \in [0,1]$; and
- $HD(X, X) = 0$.

5.2. Virtual distance and virtual similarity

The Hamming distance operation implicitly treats the pair (0, 0) as a similar pair. However, what is really concerned in this study are the non-zero, or (1, 1), pairs. In order to sort out the non-zero pairs, the virtual distance $VD(X, Y)$ is introduced

$$VD(X, Y) = 1 - \frac{\langle X \cap Y \rangle}{N} \tag{5}$$

This distance accounts for the proportion of exclusive occurrence of X or Y [the (1, 0) and (0, 1) pairs] and non-occurrence of X and Y [the (0, 0) pairs]. It is a measure of the probability of X and Y not occurring simultaneously or jointly. In other words, it is the probability of X and Y occurring mutually exclusively. Virtual distance has the following property:

- Symmetry: $VD(X, Y) = VD(Y, X)$,
- $VD(X, Y) \in [0,1]$; and
- $VD(X, X) \in [0,1]$.

The complement of $VD(X, Y)$

$$VS(X, Y) = 1 - VD(X, Y) = \frac{\langle X \cap Y \rangle}{N} \tag{6}$$

is called virtual similarity. It accounts for the proportion of non-zero similar pairs (1, 1) in the two strings. The value of $VS(X, Y)$ is interpreted as the estimate of the probability of finding a sample that possesses two characteristics X and Y simultaneously, i.e., the joint probability $P(X \cap Y)$. In the context of this study, the joint probability for simultaneous occurrence of structural fire hazards R_i and R_j is denoted by

$$P_{ij} = P(R_i \cap R_j) = VS(R_i, R_j) = \frac{\langle R_i \cap R_j \rangle}{N} \tag{7}$$

It is also noted that

$$P_{jj} = P(R_j \cap R_j) = P(R_j) = P_j \tag{8}$$

5.3. Jaccard distance

Another measure of the difference between two sets or two strings of equal length is the Jaccard distance [24]. It is similar to Hamming distance but is measured on the basis of the size of the subset of the union of the two strings, i.e., $X \cup Y$. Denote Jaccard distance by $JD(X, Y)$. It is expressed as

$$JD(X, Y) = \frac{\langle X \cup Y \rangle - \langle X \cap Y \rangle}{\langle X \cup Y \rangle} \tag{9}$$

Or

$$JD(X, Y) = 1 - \frac{\langle X \cap Y \rangle}{\langle X \cup Y \rangle} \tag{10}$$

Jaccard distance is a measure of dis-similarity between two objects. Jaccard distance can be interpreted as a conditional probability, i.e., it is the probability of finding dis-similar components among all pairs (x_i, y_i) (i=1, 2, ..., N) under the condition that at least one component in the pair attains a non-zero value. The Jaccard distance $JD(X, Y)$ has the following properties:

- Symmetry: $JD(X, Y) = JD(Y, X)$
- $JD(X, Y) \in [0,1]$; and
- $JD(X, X) = 0$.

It is noted that Hamming distance, Jaccard distance and the virtual distance have similar properties. Moreover, since $N \geq \langle X \cup Y \rangle$, it follows from Eqs. (4), (5) and (10) that

$$HD(X, Y) \leq JD(X, Y) \leq VD(X, Y) \tag{11}$$

5.4. Conditional probability of occurrence

Conditional probability has the form

$$P(Y|X) = \frac{\langle X \cap Y \rangle}{\langle X \rangle} \tag{12}$$

This is the probability of finding Y under the condition that X has occurred. Similarly, the expression

$$P(X|Y) = \frac{\langle X \cap Y \rangle}{\langle Y \rangle} \tag{13}$$

is the probability of finding X under the condition that Y has occurred.

Unlike Jaccard distance, there is no symmetry between the two conditional probabilities $P(Y|X)$ and $P(X|Y)$, i.e., generally

$$P(Y|X) \neq P(X|Y) \tag{14}$$

However, it can be shown (see Appendix C) that Bayes rule

$$P(X|Y)P(Y) = P(Y|X)P(X) \tag{15}$$

applies.

5.5. Phi correlation coefficient of fire hazard pairs

It is possible that for a given housing property or a group of housing properties, multiple fire hazards may co-exist. In order to develop a strategy to alleviate the fire hazards, one may be interested in how the fire hazards are correlated. To measure correlation between binary variables, we use the Phi correlation coefficient [25].

$$\phi_{i,j} = \frac{N_{ij} N_{\bar{i}\bar{j}} - N_{i\bar{j}} N_{\bar{i}j}}{\sqrt{N_i N_{\bar{i}} N_{\bar{j}} N_j}} \tag{16}$$

where N_{ij}, $N_{\bar{i}\bar{j}}$, $N_{i\bar{j}}$, $N_{\bar{i}j}$ is the frequency of houses that have hazard i and j, have neither hazards i and j, have hazard i but not j, and have hazard j but not i, and N_i, N_j, $N_{\bar{i}}$, $N_{\bar{j}}$ is the frequency of houses that have hazard i, have hazard j, not have hazard i, and not have hazard j, respectively. Note that Phi correlation is equivalent to Pearson correlation for binary variables having values 1 and 0.

5.6. Joint probability of the occurrence of multiple structural fire hazards

A discussion has been given in Section 5.2 on the joint probability of the occurrence of any pair of two structural fire hazards. Generally, the probability of finding a sample that possesses multiple characteristics X, Y, Z, ... simultaneously can be estimated from

$$P(X \cap Y \cap Z \cap \ldots) = \frac{\langle X \cap Y \cap Z \cap \ldots \rangle}{N} \tag{17}$$

For simplicity, the following notation is introduced for joint probability of three fire hazards

$$P_{i,j,k} = P(R_i \cap R_j \cap R_k) = \frac{\langle R_i \cap R_j \cap R_k \rangle}{N} \tag{18}$$

The joint probability has the following properties:

- Commutative: $P_{ijk} = P_{jik} = P_{ikj} = P_{(\text{any combination of } i,\ j,\ k)}$
- $P_{ijj} = P_{iij} = P_{ij}$
- $P_{iii} = P_{ii} = P_i = P(R_i) = \pi_i$

Similarly, the joint probability of four fire hazards is defined as

$$P_{i,j,k,l} = P(R_i \cap R_j \cap R_k \cap R_l) = \frac{\langle R_i \cap R_j \cap R_k \cap R_l \rangle}{N} \tag{19}$$

The aforementioned properties also apply to $P_{i,j,k,l}$.

6. Results and discussion

An algorithm written using the R programming platform was developed to perform the statistical analyses described above. The results are presented and discussed in the following subsections.

6.1. Hamming distance results

A fire hazard can be treated as an attribute of a sample. In the current study, the sample is of size 47 with 8 attributes examined. The results of the Hamming distance $HD(R_i, R_j)$ for all pairs (i, j=1, 2, ..., 8) of fire hazards are recorded in Table 3 in the form of a matrix. Because of the symmetry property, i.e., $HD(R_i, R_j)$=$HD(R_j, R_i)$ only one half of the matrix is presented.

From Table 3 we can see that the smallest Hamming distance of 0.085 is between fire hazards R_2 and R_7, indicating that the two are closely related. Referring to Table 1, this result means that the chance of finding box gutters penetrating the separating wall whist not having combustible separating walls between adjacent balconies of attached houses, and vice versa, is very low, indicating some association between the two fire hazards. The highest Hamming distance value (0.617) is between R_1 and R_2, indicating that the corresponding two fire hazards, i.e., timber penetrations through the separating wall and box gutters

Table 3
The Hamming distance results for all pairs of fire hazards.

	R_1	R_2	R_3	R_4	R_5	R_6	R_7	R_8
R_1	0	0.617	0.447	0.383	0.340	0.468	0.574	0.448
R_2		0	0.255	0.362	0.319	0.234	0.085	0.448
R_3			0	0.277	0.234	0.191	0.212	0.310
R_4				0	0.255	0.255	0.319	0.448
R_5					0	0.255	0.277	0.276
R_6						0	0.191	0.276
R_7							0	0.345
R_8								0

Table 4
The Jaccard distance results for all pairs of fire hazards.

	R_1	R_2	R_3	R_4	R_5	R_6	R_7	R_8
R_1	0	1	0.875	0.72	0.696	0.917	1	0.765
R_2		0	1	1	1	1	0.500	1
R_3			0	0.927	0.917	1	1	0.900
R_4				0	0.800	0.924	1	1
R_5					0	1	1	1
R_6						0	1	0.889
R_7							0	0.909
R_8								0

penetrating the separating wall, are least associated with each other.

Note that Hamming distance includes simultaneous non-occurrence as well as simultaneous occurrence of hazards pairs. The high value of $HD(R_2, R_7)$ means that both the occurrence and non-occurrence patterns of the two hazards are similar, it does not necessarily mean that the relative frequency for joint occurrence of hazards R_2 and R_7 is high. The discussion on probability for joint occurrence of paired hazards is given in later in virtual similarity analysis.

6.2. Jaccard distance results

The results of applying the Jaccard distance to pairs of fire hazards is presented in Table 4. Again, only one half of the matrix is presented because of symmetry.

The dis-similarity between fire hazards can be estimated from Table 4. A value of 1 indicates that the two hazards are mutually exclusive, whereas a value of 0 indicates that the pair of fire hazards only appeared together. The most associated fire hazards are R_2 and R_7, with the lowest JD value of 0.5. This result indicates that once hazard R_2 is detected, there is a good chance that hazard R_7 will also be detected, and vice versa. This result is in concurrence with that of Hamming distance analysis. The Jaccard distance value of 1 for a given pair of (R_i, R_j) means that the pair are mutually exclusive, i.e., once hazard R_i is detected, R_j will not be, and vice versa. It can be seen from Table 4 that hazard R_2 is mutually exclusive against all other hazards except for R_7. Comparing Tables 3 and 4, it is seen that the inequality as in Eq. (11) is satisfied.

6.3. Virtual similarity, joint probability and conditional probability

The virtual pairwise similarities, or the joint probabilities $P(R_i \cap R_j)$ (i, j=1, 2, ..., 8) of the pairwise simultaneous occurrence of fire hazards are presented in Table 5. Because of the symmetry, only one half of the table is presented. Since $P(R_i \cap R_i) = P(R_i)$, the diagonal probabilities in Table 5 are the probabilities, or the frequencies of occurrence of a given hazard listed in Table 1.

It is seen from Table 5 that hazard R_2 does not co-exist with other hazards except for R_7, though the probability for their joint occurrence is small (8.5%). It can be deduced that the probability for R_2 to be the only hazard in a heritage housing property is

$$P(R_2) - P(R_2 \cap R_7) = 0.149 - 0.085 = 0.064 \quad (20)$$

and that the probability for R_7 to be the only hazard in a heritage housing property is

$$P(R_7) - P(R_2 \cap R_7) - P(R_8 \cap R_7) = 0.106 - 0.085 - 0.021 = 0 \quad (21)$$

The pairs of events that have the highest probability of co-occurrence are (R_1, R_4) and (R_1, R_5).

The pairwise conditional probabilities $P(R_i \mid R_j)$ for i, j = 1, 2, ..., 8 are presented in Table 6.

The results have shown that $P(R_1 \mid R_5)$ attains the highest value (0.875) among all estimated conditional probabilities. This means that once a skylight was added to a heritage housing property at a location that is within 900 mm of the adjoining property, the chance of finding timber penetrations through the wall separating the two properties is high. However, the events in the opposite order, i.e., detecting a non-compliant skylight once timber penetration has been found, has a considerably lower probability [$P(R_5 \mid R_1) = 0.318$]. The conditional probabilities $P(R_2|R_7)$ and $P(R_7|R_2)$ are the second and third highest, indicating that the two structural fire hazards are closely related. This result is in agreement with the outcome of the foregoing Hamming distance and Jaccard distance analyses. In other words, the structural fire hazards of box gutters penetrating the separating wall and having combustible separating walls between adjacent balconies of attached houses are closely related, though the probability of occurrence for either individual hazards is not necessarily high (see Table 1). Furthermore, despite the chance for their simultaneous occurrence is small (see Table 5), the probability of finding R_7 once R_2 is detected is relatively high (0.571) and vice versa (0.8). Similarly, there is a good chance of finding R_4 or R_5 once R_1 is detected (both being 0.318) and vice versa (0.7 or 0.875 respectively).

Table 5
Pairwise joint probability $P(R_i \cap R_j)$.

i	j							
	1	2	3	4	5	6	7	8
1	0.468	0	0.064	0.149	0.149	0.043	0	0.043
2		0.149	0	0	0	0	0.085	0
3			0.106	0.021	0.021	0	0	0.021
4				0.213	0.064	0.021	0	0
5					0.170	0	0	0.021
6						0.085	0	0.021
7							0.106	0.021
8								0.149

Table 6
Pairwise conditional probability $P(R_i \mid R_j)$.

i	j							
	1	2	3	4	5	6	7	8
1	1.000	0.000	0.600	0.700	0.875	0.500	0.000	0.286
2	0.000	1.000	0.000	0.000	0.000	0.000	0.800	0.000
3	0.136	0.000	1.000	0.100	0.125	0.000	0.000	0.143
4	0.318	0.000	0.200	1.000	0.375	0.250	0.000	0.000
5	0.318	0.000	0.200	0.300	1.000	0.000	0.000	0.143
6	0.091	0.000	0.000	0.100	0.000	1.000	0.000	0.143
7	0.000	0.571	0.000	0.000	0.000	0.000	1.000	0.143
8	0.091	0.000	0.200	0.000	0.125	0.250	0.200	1.000

6.4. Result of Phi correlation analysis

The correlation coefficients between all pairs of fire hazards are presented in Table 7. Because of symmetry, i.e., $\phi_{i,j}=\phi_{j,i}$, only half of the data matrix is presented. It can be seen in Table 7 that the greatest correlation coefficient (0.631) is between $\boldsymbol{R_2}$ and $\boldsymbol{R_7}$. The next largest correlation is an anti-correlation (−0.392) between $\boldsymbol{R_1}$ and $\boldsymbol{R_2}$, meaning that when one occurs, the other is not likely to occur. Fire hazard $\boldsymbol{R_1}$ is also correlated with $\boldsymbol{R_5}$ and anti-correlated with $\boldsymbol{R_7}$. Hazard R_2 is mutually exclusive to all other hazards except for R_7. This result agrees well with that of Jaccard distance analyses (see Table 4 and the corresponding discussions). It can also be seen from Table 7 that the fire hazards $\boldsymbol{R_3}$ and $\boldsymbol{R_6}$ have little correlation with the other fire hazards. All the above observations agree well with Jaccard distance, virtual similarity and conditional probability analyses.

To visualise the relationship between each pair of fire hazards, the Multidimensional Scaling approach is undertaken. Multidimensional scaling takes a matrix of distances between all pairs of items in a set, and projects the items onto a two dimensional space [26]. The two

Table 7
Pairwise sample correlation $\phi_{i,j}$ between each pair of fire hazards.

	R_1	R_2	R_3	R_4	R_5	R_6	R_7	R_8
R_1	1.000	−0.392	0.091	0.242	0.369	0.020	−0.324	−0.129
R_2		1.000	−0.144	−0.217	−0.189	−0.128	0.631	−0.216
R_3			1.000	−0.011	0.027	−0.105	−0.119	0.015
R_4				1.000	0.180	0.028	−0.179	−0.304
R_5					1.000	−0.138	−0.156	−0.079
R_6						1.000	−0.105	0.169
R_7							1.000	−0.035
R_8								1.000

Table 8
The estimated non-zero 3-tuple hazards joint probabilities $P_{i,j,k}$.

i	j	k	$P_{i,j,k}$
1	3	5	0.021
1	3	8	0.021
1	4	5	0.064
1	4	6	0.021
1	5	8	0.021
1	6	8	0.021
3	5	8	0.021

dimensional projection obtained is one that best preserves the original pairwise distances between each fire hazard. By doing this, the relationships between the set of high dimensional elements can be visualised in a two dimensional plot. To visualise the correlation relationship between each of the fire hazards, the correlations are first mapped to distances using the distance function $D = 1 - |C|$, where C is the correlation matrix and D is the distance matrix. Two fire hazards are related if they are correlated or anticorrelated; the absolute value of C is used to reflect this in the distance function. The projection of the distances between the eight structural fire hazards is shown in Fig. 3.

Note that the plot is a projection of the 47 dimensional space into a two dimensional space. The original space contained one dimension per sample (house); the projected space contains two dimensions that are a linear combination of the correlation distances. The labelling of these dimensions is not important. It is the relative distances between each of the projected points that are important.

It can be seen in this figure that fire hazards R_2 and R_7 show a strong relationship, followed by R_5 and R_1. Hazards R_4 and R_8, and R_6 and R_3 show a weak relationship. More specifically for the studied heritage building stock in Sydney, it appears that the following pairs of fire hazards are more likely to occur together than any other pairs:

- 'box gutters penetrating the separating wall' and 'combustible separating walls between adjacent balconies of attached houses';
- 'separating wall stopping short of the roof' and 'skylights installed within 900 mm of the adjoining property';
- 'timber penetrations through the separating wall' and 'skylights installed within 900 mm of the adjoining property'.

The reasons for the relatively high joint occurrences of the above listed pair of hazards are attributed to the nature of the renovation works. For example, adding a skylight in the roof is likely to require extra timber beam or truss, which may result in timber penetrations through the separating wall.

Now that the correlations between the fire hazards have been unveiled, the significance of the correlations needs to be tested. Unfortunately, it is unsure if the correlation is from a direct or spurious

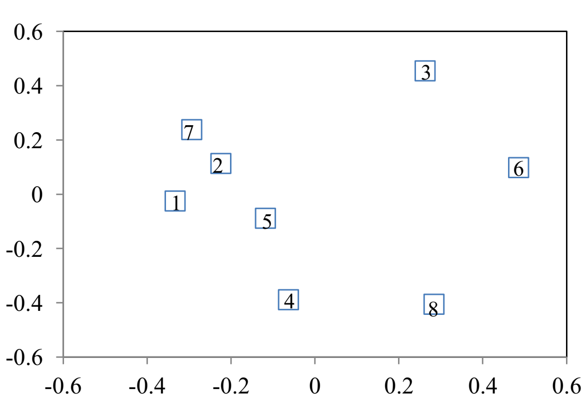

Fig. 3. Mapping of correlation distances between pairs of fire hazards.

relationship. For example, the fire hazard X might be correlated with fire hazards Y and Z, but this may be due to X being dependent on Z and Y being dependent on Z. Therefore, knowledge of the presence or absence of fire hazard Z is sufficient to determine the probability of X, meaning Y is not required. The relationships between various fire hazards will be examined further in next subsection through joint probability analysis and in Section 7 through logistic regression modelling to identity direct relationships between fire hazards.

6.5. Result of joint probability of the occurrence of more than two multiple structural fire hazards

The evaluations of the third order joint probabilities revealed that most $P_{i,j,k}$ attain a value of zero. The non-zero values of $P_{i,j,k}$ for the sampled buildings are listed in Table 8.

Some statements can be made about this table: fire hazard R_1 is the most popular hazard that is likely to co-exist with at least 2 other hazards; the second most popular hazard is R_8 followed by R_3. In particular, hazards R_1, R_4 and R_5 are the most likely combination of co-existence. The handful of small non-zero values of $P_{i,j,k}$ is consistent with the small probability value of $1-C_2$, or the small probability of finding more than 2, or at least 3 structural fire hazards in a heritage housing property simultaneously (see Fig. 2). Note that the entries to Table 8 are not limited to samples that have 3 hazards only. For example, a building that possesses m ($m\geq3$) hazards will contribute $\binom{3}{m} = \frac{m!}{3!(m-3)!}$ entries to this table. In fact, it is property number 37 (see Table A1) that contributed the most to the entries to Table 8, because this property possessed 4 hazards.

The non-zero result of the forth order joint probability $P_{i,j,k,l}$ for the sampled buildings are listed in Table 9.

According to Table 9, there is only one possible co-existence of four fire hazards with probability of 0.021. This result is consistent with the probability value of $1-C_3$ given in Fig. 2. Again the property that contributed to the sole entry is property number 37. From Table 8 and Table 9 it can be concluded that although hazard R_8, or the hazard of having common eaves with adjoining properties, does not occur frequently, when it does, it is likely to be associated with two or more other hazards.

Since the probability for a property to have more than 4 structural fire hazards in the investigated heritage housing properties is zero (see Fig. 2), higher than 4-tuple hazards joint probabilities are all zero.

7. Identifying direct relationships

It is seen in the previous section that many of the fire hazards are highly correlated, meaning that there could be dependence between

Table 9
Estimated non-zero 4-tuple hazards joint probability $P_{i,j,k,l}$.

i	j	k	l	$P_{i,j,k,l}$
1	3	5	8	0.021

them. However, the pairwise correlations may be either direct or indirect due to a confounding factor. In this section the concept of indirect correlation is examined through the use of the logistic regression modelling.

7.1. Spurious relationships

Given that the correlation between two fire hazards R_i and R_j has been identified (i, j=1, 2, ..., M, and i‡j), the explanation for the correlation may be:

- the correlation is coincidence and the two variables are independent,
- the occurrence of fire hazard R_i depends on the appearance of fire hazard R_j,
- the appearance of fire hazard R_j depends on the appearance of fire hazard R_i,
- the appearance of fire hazards R_i and R_j depend on the appearence of a third confounding fire hazard R_k. (k‡i and k‡j)

The first three cases describe the relationship between the two variables R_i and R_j as either absent (correlation is a coincidence), left to right (R_i effects R_j) or right to left (R_j effects R_i). The fourth case introduces another fire hazard R_k in which both R_i and R_j are dependent on, giving the appearance of dependence between R_i and R_j.

Because of the possibilities of co-existence of more than two or three fire hazards, multiple confounding fire hazards may exist. It is difficult to reveal any complex relationships between various fire hazards by examining the pairwise correlations alone. A method to avoid any confounding fire hazards or to reveal the complex relationship between various fire hazards is to model a given fire hazard variable with respect to all other variables.

7.2. Regression modelling of the fire hazards

The study in Section 5 examined the pairwise relationships between any two of the eight identified structural fire hazards through pairwise joint probabilities. It is highly possible that more complex relationships exist between various hazards. In this section, the correlation is examined in more detail by modelling the correlation of each fire hazard amongst the remaining fire hazards.

To model the relationship between each fire hazard to the remaining fire hazards, a function of the form is assumed:

$$R_i = f(R_{-i}) \tag{22}$$

where R_i is the indicator variable representing the presence of fire hazard i, R_{-i} is the set of indicator variables representing the remaining fire hazards, and $f(\bullet)$ is a function casting the relationship between R_{-i} and R_i.

The simplest model is a linear model of the form:

$$\hat{R}_i = \sum_{j\neq i}\beta_j R_j + \beta_0 \tag{23}$$

where $\hat{R}_i$ is the predicted value of the response indicator variable, R_j (where $j\neq i$) are the indicator covariates that are being examined, β_j are the coefficients of the model that depict the correlation between R_i and R_j, and β_0 is the model offset (the expected value of R_i when all other fire hazards are absent).

Unfortunately, this form of model treats the predicted response ($\hat{R}_i$) as an element from the real domain $(-\infty, +\infty)$, but our response variables are indicator variables that can only take the values [0,1] where 0 represents that the fire hazard is absent and 1 represents the fire hazard being present.

A more suitable model is a logistic regression, which has the form:

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \sum_{j\neq i}\beta_j R_j + \beta_0 \tag{24}$$

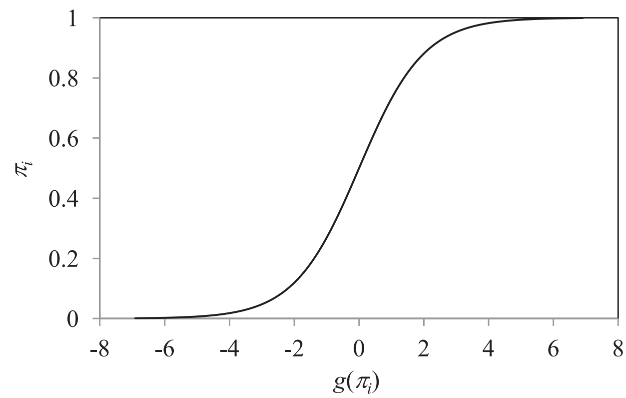

Fig. 4. Logistic link function.

where π_i is the estimated probability of fire hazard i being present. The right hand side of the equation still has the linear form, but the left hand side has transformed the response using the logistic link function. The logistic link function transforms the probability domain [0,1] to the real domain $(-\infty,+\infty)$ allowing us to fit a probability to a linear equation. The transformation converts the probability of the event π_i to the odds of the event $\pi_i/(1-\pi_i)$, then to the log odds of the event. The logistic link function:

$$g(\pi_i) = \log\left(\frac{\pi_i}{1-\pi_i}\right) \tag{25}$$

is presented in Fig. 4. It can be seen that the logistic link function maps the probability of 0.5 to the real value 0 and as the probability approaches 1, the associated real value approaches $+\infty$. Also, as the probability approaches 0, the associated real value approaches $-\infty$.

In a simple linear regression of the form $\pi_i = \sum_{j\neq i}\beta_j R_j + \beta_0$, if R_j changes from 0 to 1, the response in R_i increases by β_j. However, when using logistic regression as in Eq. (24), it is not obvious how the change in variables effects the response. The behaviour of the logistic regression is examined in the remainder of this section.

If all covariate fire hazards are zero (absent), then the model is expressed as:

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 \tag{26}$$

showing that e^{β_0} is the odds of fire hazard R_i being present, or $\pi_i = (1+e^{-\beta_0})^{-1}$, given that all others are not present.

Comparing the change in odds of R_i when fire hazard R_j changes from absent to present, it can be found that $g(\pi_i)$ increases by β_j.

$$\begin{aligned}\log\left(\frac{\pi_i}{1-\pi_i}\right) &\rightarrow \log\left(\frac{\pi_i}{1-\pi_i}\right) + \beta_j\\ &= \log\left(\frac{\pi_i}{1-\pi_i}\right) + \log[\exp(\beta_j)]\\ &= \log\left[\frac{\pi_i}{1-\pi_i}\exp(\beta_j)\right]\end{aligned} \tag{27}$$

Therefore the appearance of risk factor R_j has increased the odds of risk factor R_i appearing by $\exp(\beta_j)$. This means that if β_j is zero, the risk factor R_j has no effect on risk factor R_i. If β_j is positive, the risk factor R_j increases the probability of R_i, and if β_j is negative, the risk factor R_j decreases the probability of R_i.

The process as expressed by Eq. (27) can be extended to include all R_j (j‡i) and the resulting the model can be rearranged to take the following form:

$$\pi_i = \frac{1}{1+\exp\left(-\sum_{j\neq i}\beta_j R_j - \beta_0\right)} \tag{28}$$

allowing the computation of the estimated probability of the fire hazard R_i.

7.3. Fitting the model

In order to compute suitable β_i coefficients, the logistic regression model is fitted to the available data. The most suitable coefficients for the data are found by choosing the set of β_i values that maximise the likelihood of the data. The likelihood of the data is obtained by computing the probability of the data occurring, or the multiplication of the probability of all of the events. So to determine the likelihood of fire hazard R_i, the likelihood function of the form:

$$L(R_i) = \prod_{k=1}^{N} \pi_{i,k}^{R_{i,k}} (1-\pi_{i,k})^{(1-R_{i,k})} \tag{29}$$

is evaluated, where $\pi_{i,k}$ is the probability of fire hazard R_i in the kth sampled building and $1 - \pi_{i,k}$ is the probability of the hazard not occurring, $R_{i,k}$ is 1 if the hazard is present in building k and 0 if the hazard is absent, and N is the sample size. Substituting the logistic regression function into the likelihood function, the coefficients β_i that provide the maximum likelihood can be obtained by setting the gradient of the likelihood function to zero and solving the resultant set of equations. Further details of this process can be found in Hastie et al. [27].

Once the model has been established, the coefficients are examined to identify how the set of fire hazards effect each other. The fire hazard R_i can be treated as having a mean and variance. The mean is the expected value when no other information is given. The variance is related to how different the truth is from the mean. If the model only has the offset coefficient β_0, it can only model the mean. As more coefficients are introduced into the model (using information about other fire hazards R_j), more variance of the response can be explained. So as more coefficients are introduced, the likelihood will either remain the same or increase, but never decrease.

The objective is to determine which subset of fire hazards effect a given fire hazard R_i, meaning that a function is needed to provide a measure of the suitability of the model for the data. If the likelihood was used as a measure of suitability, the complete model for all fire hazards (each fire hazard is a function of all other fire hazards) would be derived. However, the complete model is not necessarily the most suitable model, since confounding factors or indirect correlations may also be included.

A better measure of suitability is the Akaike information criterion (AIC) [28]. This measure trades off between the likelihood of the data and the number of terms in the model. Let m denote the number of terms included in the model and L_m the corresponding likelihood evaluated from the data for the given model. AIC is defined as:

$$\text{AIC} = 2m - 2\log(L_m) \tag{30}$$

Note that since L_m is always less than 1, AIC is a non-negative function. The Akaike information criterion states that the smaller the AIC value is, the more suitable is the model. According to Eq. (30), the decrease in $-2\log(L_m)$, due to the increase in L_m by adding more terms, may not necessarily result in a decrease in AIC because 2 m is increased. On the other hand, drastically reducing m may result in significant reduction in L_m, hence an increase in AIC. An optimum value of m may exist such that the corresponding AIC function attains a minimum.

To determine the most suitable model for each fire hazard R_i, a stepwise search method is used, whereby the AIC function of the full model (i.e., π_i as a function of all other $M-1$ fire hazards, M=8 for the current study) is initially computed. Then the AIC values of the model with one less term (π_i as a function of $M-2$ fire hazards) are computed to obtain $M-1$ new AIC values. If the smallest of these $M-1$ AIC values is less than the full model AIC, the corresponding model is selected for further search of a better model with 2 less terms (or $M-2$ terms). This process is repeated until the AIC value does not decrease and the resultant model is determined to be the final model.

Table 10
Complete model terms and the AIC value for R_1.

Model Terms	AIC
R_2, R_3, R_4, R_5, R_6, R_7, R_8	44.8

Table 11
Combinations of one less term and the corresponding AIC values for R_1.

Model Terms	AIC
R_3, R_4, R_5, R_6, R_7, R_8	42.8
R_2, R_4, R_5, R_6, R_7, R_8	42.8
R_2, R_3, R_5, R_6, R_7, R_8	43.6
R_2, R_3, R_4, R_6, R_7, R_8	46.1
R_2, R_3, R_4, R_5, R_7, R_8	43.2
R_2, R_3, R_4, R_5, R_6, R_8	43.2
R_2, R_3, R_4, R_5, R_6, R_7	43.0

For example, the full model terms and the computed AIC value for R_1 are given in Table 10.

When removing one term, the model terms and the corresponding AIC values as listed in Table 11 are obtained.

The top two rows of combinations in Table 11 yield the least value of AIC, indicating that a suitable model is obtained when R_2 or R_3 is removed. After R_2 is removed (or β_2 is set to zero), this process is repeated with the new model to remove other variables until no further reduction in the AIC is achievable.

Finally a step of bias-reduction is required due to the small sample size. For a given set of coefficients, if the response has probability 0 or 1, the linear sum is mapped to $+\infty$ or $-\infty$, and provides us with overestimates of the coefficients β_i. Therefore, once the terms have been finalised using above stepwise AIC process, the model is then refitted using Bias-Reduced Logistic Regression [29] to provide reasonable coefficient values. Bias-reduced logistic regression introduces a bias into the response probability to ensure that probabilities of 0 or 1 are not obtained.

The above process was conducted for all eight identified structural fire hazards and the final models for each of the hazards are presented in Table 12. The second column of the table contains the chosen non-zero coefficients for the model (where β_0 is the model intercept and β_i is associated to R_i; the coefficients not listed in the table were removed by AIC and are set to zero in the model), the third column contains the estimated value for the coefficient, the forth column contains the standard error of the coefficient, the fifth column contains the z-value test statistic, and the sixth column contains the associated p-value. Note that the standard error, z-value and resulting p-value are computed under the assumption that no other models have been tested. We have already tested a set of models using AIC and arrived at the presented models, therefore these values are less insightful for the final models, but are included in Table 12 for completeness.

The estimate of coefficient β_0 gives us information about the probability π_i when each of the fire hazard variables R_i=0. The estimate of β_j provides us with the effect of the presence of R_j on π_i. If β_j is zero, then R_j has no effect on π_i. If β_j is positive, then the presence of R_j increases the probability of π_i. If β_j is negative, then the presence of R_j decreases the probability of π_i.

The standard error of each β_j gives us an indication of the variability of the coefficient across the data. If the standard error is low, then the estimated value is a good estimate for the whole data set. If the standard error is high, then the coefficient is highly variable and the estimate may not be useful.

The z-value is the test statistic for the hypothesis test where the Null Hypothesis is that the true value of β_i is zero, and the alternative is that

Table 12
Estimated model coefficients and statistics.

Dependent Variable	Coefficients	Estimate	Std. Error	z-value	p-value
R_1	β_0	−0.1214	0.3542	−0.343	0.7319
	β_2	−2.5867	1.6011	−1.616	0.1062
	β_5	1.7308	1.0127	1.709	0.0874
R_2	β_0	−3.628	1.436	−2.527	0.01151
	β_7	5.788	2.156	2.685	0.00725
	β_8	−3.216	2.747	−1.171	0.24161
R_3	β_0	−2.0448	0.4579	−4.465	8.01e−06
R_4	β_0	−0.5108	0.4739	−1.078	0.281
	β_2	−1.6864	1.7327	−0.973	0.330
	β_8	−2.1972	1.6318	−1.347	0.178
R_5	β_0	−2.7932	0.8578	−3.256	0.00113
	β_1	2.0673	0.9709	2.129	0.03323
R_6	β_0	−2.2687	0.5004	−4.533	5.8e−06
R_7	β_0	−3.664	1.469	−2.493	0.01266
	β_2	5.861	2.222	2.638	0.00835
	β_8	2.197	1.760	1.249	0.21183
R_8	β_0	−0.6539	0.4968	−1.316	0.188
	β_2	−3.2958	2.8480	−1.157	0.247
	β_4	−2.0541	1.6386	−1.254	0.210
	β_7	1.7525	2.3622	0.742	0.458

the true value of β_i is not zero. The greater the absolute value of the test statistic, the less likely that β_i is zero. The associated p-value provides us with the probability that the true value of $\beta_i = 0$. Note that a large p-value (say greater than 0.1) does not mean that the true corresponding value of β_i is zero, it only means that there is insufficient support from the data to show that it is not zero.

The results show that $\boldsymbol{R_3}$ and $\boldsymbol{R_6}$ are not effectively affected by any of the other fire hazards. This result agrees with that of the correlation analysis as shown in Table 7 where $R_{3,j}$ ($j\neq3$) and $R_{6,j}$ ($j\neq6$) are relatively small. The intercept provides the expected probability of $\boldsymbol{\pi_3}$ and $\boldsymbol{\pi_6}$ [$\pi_3 = (1 + e^{2.044})^{-1} = 0.1146$ and $\pi_6 = (1 + e^{2.268})^{-1} = 0.0938$] that agree with the results presented in Table 1.

If the relationships at the 10% level of significance (meaning that there is less than a 10% chance that the correlation is zero) is examined, R_1 is seen to positively correlated with R_5, R_2 is highly positively correlated with R_7, R_5 is positively correlated with R_1, and R_7 is highly positively correlated with R_2. It is interesting to find that none of the correlations with p-value less that 0.1 were negative. Note that although R_8 appears as a correlated variable in three of the models, the associated p-value is greater than 0.1, implying that the relationships are not significant. Note that these p-values have been computed after model selection using AIC and therefore are likely to contain bias [30]. R_1 is shown to have the greatest probability of occurrence (shown by it having the largest β_0 of all fitted models), but it is found that it has little use in predicting the state of many of the risk factors. The model for R_5 is the only model that is shown as a function of R_1. The similarity between R_1 and R_5 can also be seen in Fig. 3.

Further comparison between Fig. 3 and Table 12 shows that some of the variables are modelled on their close neighbours in Fig. 3(e.g. R_2 is a function of R_7 and R_8, and R_5 is a function of R_1), but some are modelled on distant neighbours (e.g. R_2 is a function of R_7 and R_8, and R_5 is a function of R_1). It must be remembered that the models are constructed using the variables that best explain the variance of the response variable, so for the cases where fire hazards have been modelled using distant neighbours, it is found that the negative correlation from these distant neighbours are better predictors than the positive correlation from the close neighbours. It can also be seen that for each model with distant neighbours, the associated coefficient for the distant neighbour is negative, implying that when the fire hazard is present, the probability of the response fire hazard decreases.

To examine the above individual models, each model's change in deviance from the full model is also computed. It is found that the deviance of each model (a measure of dis-similarity from the best possible fit) has no statistically significant difference from the full model containing all seven coefficients. This means that the probability estimates provided by the above models are just as accurate as the probability estimates provided by the full models, for the dataset. Therefore, by reducing the models, no information has been lost.

Regression analysis using logistic regression and AIC based step-wise variable selection revealed that R_3 and R_6 have no significant association to the other fire hazards. In other words, the construction and refurbishment events that have led to the individual structural non-compliances to the building code in respect to R_3 and R_6 are unlikely to cause other structural fire hazards. On the other hand, R_1 and R_5 have a symmetric association. Referring to Table 1, it is not hard to find the explanation in reality, since it is highly possible that “timber penetrations through the separating wall” was part of the renovation to install skylights “within 900 mm of the adjoining property”. Similarly, the symmetric association between R_2 and R_7 gives a statistical evidence that when the design and installation of a building feature, i.e., the walls separating adjacent balconies, are non-compliant to the building regulations, the materials used for the feature are also likely to be non-compliant. These results are consistent with the correlation mapping in Fig. 3 and with the analysis of the data in Table 6 (see discussion in Section 6.3). The regression analysis also showed that presence of a fire hazard does not decrease the probability of another type of fire hazard.

8. Conclusion

The concepts of Hamming distance and Jaccard distance, which are used in digital signal analysis, have been given statistical or probabilistic interpretations for fire hazard analysis. They, together with the introduced virtual distance concept, were employed in the current study to obtain estimates of the association, joint probabilities and conditional probabilities for the occurrence of identified structural fire hazards from a sample group of heritage housing properties in Sydney. The analyses revealed interesting pairwise relationships between various hazards of the sampled heritage housing properties. In addition, Phi correlation and regression analyses were conducted to uncover the relationship between the identified structural fire hazards. The results by the various methods are consistent.

It has been shown that the probability of finding at least one structural fire hazard in heritage housing stock in Sydney is as high as 87%. The probability of a heritage building possessing multiple structural fire hazards is also high (44%). Such significant values indicate high likelihood of structural damage and fire spread to adjacent properties in events of fires, and non-negligible risk to heritage housing protection. When examining the joint probability between fire hazards, it is found that there is a small chance that more than three fire hazards will occur.

The joint and conditional probability analyses showed that some structural fire hazard in the sampled group of heritage housing properties are related and the correlations are confirmed by Phi correlation and logistic regression analyses. In particular, it has been observed that a hazardous structural feature as a result of renovation is likely to be supported by another hazardous feature. Evidence also exists that when the design and installation of a building feature are non-compliant to building regulations, the materials used for the feature are also likely to be non-compliant.

The quantified results of structural fire hazard occurrence probabilities, the significant joint probabilities and associations provide support to the recommendations in the previous study that authorities having jurisdictions should consider the introduction of compulsory building surveying audit wherever a refurbishment is undertaken in a heritage housing property regardless of the extent of the refurbishment. The outcome of this study also provide guidance for fire safety engineering practice to develop effective solutions to address correlated structural fire hazards in heritage buildings.

Statistical analysis of the fire hazard occurrence frequencies obtained in a previous study involved significant ranges of uncertainty at the 95% confidence. It would be prudent to increase the sample size in order to obtain a better estimate of the fire hazard occurrence frequencies. Larger sample size is also needed to consolidate the findings of the current study.

The methods employed in the current study can be extended to the studies of structural fire hazards in any type of buildings or building classes and to the studies of other types of fire hazards. The findings can be used to guide the inspection or audit works by building surveyors or to assist fire safety engineers in risk assessment and developing cost-effective fire safety measures that help alleviate multiple fire hazards. Last, but not the least, the results can also be used by fire services to develop risk based or risk informed approach to firefighting and personnel protection in case of fire incidences.

Acknowledgement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The authors are indebted to Mr Mel Green and Dr Mary Hardie for providing the raw data of the building surveying results. Mr Jiaxin Fang's assistance in the initial R programming is also acknowledged.

Appendix A

See Table A1.

Table A1
Digitised fire hazard registration matrix.

Sample ID	Fire hazard							
	R_1	R_2	R_3	R_4	R_5	R_6	R_7	R_8
1	1	0	0	1	0	1	0	0
2	0	0	0	0	0	1	0	0
3	0	0	0	1	0	0	0	0
4	0	0	1	1	0	0	0	0
5	0	1	0	0	0	0	1	0
6	0	1	0	0	0	0	1	0
7	0	1	0	0	0	0	1	0
8	0	1	0	0	0	0	0	0
9	0	0	0	0	0	0	0	1
10	0	0	0	0	0	0	0	1
11	0	0	0	0	0	0	0	1
12	0	0	0	0	0	0	1	1
13	0	1	0	0	0	0	0	0
14	0	1	0	0	0	0	1	0
15	0	1	0	0	0	0	0	0
16	1	0	0	0	0	0	0	0
17	1	0	0	1	0	0	0	0
18	0	0	0	0	0	1	0	0
19	1	0	0	0	0	0	0	0
20	1	0	1	0	0	0	0	0
21	1	0	0	0	1	0	0	0
22	1	0	0	0	0	0	0	0
23	0	0	0	0	0	0	0	0
24	0	0	0	0	0	0	0	0
25	0	0	0	0	0	0	0	0
26	1	0	0	0	0	0	0	0
27	1	0	0	0	0	0	0	0
28	0	0	0	0	0	0	0	0
29	0	0	0	0	1	0	0	0
30	0	0	0	0	0	0	0	0
31	1	0	0	1	0	0	0	0
32	0	0	0	1	0	0	0	0
33	0	0	0	0	0	0	0	0
34	0	0	1	0	0	0	0	0
35	0	0	0	0	0	0	0	1
36	1	0	0	0	0	0	0	0
37	1	0	1	0	1	0	0	1
38	1	0	0	0	0	1	0	1
39	1	0	1	0	0	0	0	0
40	1	0	0	0	0	0	0	0
41	1	0	0	0	1	0	0	0
42	1	0	0	0	0	0	0	0
43	1	0	0	1	0	0	0	0
44	1	0	0	0	1	0	0	0
45	1	0	0	1	1	0	0	0
46	1	0	0	1	1	0	0	0
47	1	0	0	1	1	0	0	0

Appendix B

The elements x_i and y_i (i=1, 2, ..., N) of two binary strings (or sets) of X and Y with equal length take on the value of either 1 or 0. The union of the two strings

$$Z = X \cup Y \tag{31}$$

forms a new string of which each element z_i is the result of logical OR operation

$$z_i = x_i \mathrm{OR} y_i \tag{32}$$

The intersection of the two strings

$$Z = X \cap Y \tag{33}$$

is the result of logical AND operation

$$z_i = x_i \mathrm{AND} y_i \tag{34}$$

Appendix C

Without losing generally, all estimates of probabilities are referred as probabilities. In the current study, the string can be any of the column arrays in Table A1. The estimated relative occurrence frequency, or the probability of a fire hazard X is evaluated from

$$P(X) = \frac{\langle X \rangle}{N} \tag{35}$$

Likewise, the probability for the occurrence of another fire hazard Y is expresses as

$$P(Y) = \frac{\langle Y \rangle}{N} \tag{36}$$

Divide Eq. (13) by Eq. (12), we have

$$\frac{P(XY)}{P(YX)} = \frac{\langle X \rangle}{\langle Y \rangle} = \frac{\langle X \rangle / N}{\langle Y \rangle / N} = \frac{P(X)}{P(Y)} \tag{37}$$

Hence

$$P(XY)P(Y) = P(YX)P(X) \tag{15}$$

References

[1] R. McDonald, Introduction to natural and man-made disasters and their effects on buildings, Architectural Press, Oxford, UK, 2003.

[2] Geneva Association, World Fire Statistics, Geneva Association Information Newsletter, No. 27, Geneva Association, CH-1208 Geneva, ⟨https://www.genevaassociation.org/media/186303/ga2011-fire27.pdf⟩ (accessed 30 July 2016), 2011.

[3] R.W. Stevenson, Big fire in Windsor Castle raises fear about artwork, New York Times, 1992.

[4] A. Dionísio, Stone decay induced by fire on historic buildings: the case of the cloister of Lisbon Cathedral (Portugal), Geological Society 271, Special Publications, London, 2007, pp. 87–98. http://dx.doi.org/10.1144/GSL.SP.2007.271.01.10.

[5] S. Gillhuber, G. Leherberger, J. Göske, Fire damage of trachyte: investigations of the Teplá monastery building stones, Geological Society 333, Special Publications, London, 2010, pp. 73–79. http://dx.doi.org/10.1144/SP333.7.

[6] R. Shipley, S. Utz, M. Parsons, Does Adaptive reuse pay? A study of the business of building Renovation in Ontario, Canada, Int. J. Herit. Stud. 12 (6) (2006) 505–520.

[7] U. Hassler, Long-term building stock survival and intergenerational management: the role of institutional regimes, Build. Res. Inf. 37 (5) (2009) 552–568.

[8] B. Turner, TheAustralian Terrace House, Angus & Robertson, Sydney, 1995.

[9] M. Hardie, M. Green, Y. He, Fire safety and heritage protection of public housing stock in Australia, J. Cult. Herit. Manag. Sustain. Dev. 4 (2) (2014) 196–212. http://dx.doi.org/10.1108/JCHMSD-08-2013-0040.

[10] ABCB, Natinoal Construction Code of Australia, Volume 1 and 2: Building Code of Australia. Australian Building Codes Board, Canberra, Australia, 2015.

[11] J.M. Watts Jr., J.R. Hall Jr., Introduction to fire risk analysis, in: M.J. Hurley (Ed.) SFPE Handbook of Fire Protection Engineering5th ed., Society of Fire Protection Engineers, New York, 2016, pp. 2817–2826.

[12] J. Xin, C. Huang, Fire risk analysis of residential buildings based on scenario clusters and its application in fire risk management, Fire Saf. J. 62 (2013) 72–78.

[13] A. Asgary, A.S. Naini, J. Levy, Modeling the risk of structural fire incidents using a self-organizing map, Fire Saf. J. 49 (2012) 1–9.

[14] A. Anderson, O.A. Ezekoye, Property Risk Optimization by Predictive Hazard Evaluation Tool (PROPHET) - Final Report. Society of Fire Protection Engineers, ⟨https://custom.cvent.com/FDF9B074985740F68E3E55B5B684062D/files/10c412b0af2c49f98db663b7995a8723.pdf⟩, (accessed 30 July 2016), 2014.

[15] A. Silva, L.C. Neves, P.L. Gaspar, J. De Brito, Probabilistic transition of condition: render facades, Build. Res. Inf. 44 (3) (2016) 301–318.

[16] Watts, J.M.Jr., Performance-Based Fire Safety Codes for Heritage Buildings, in: R. Jonsson, K. Almand, J. Maskas, (Eds.), Proceedings of the 3rd International Conference on Performance-Based Codes and Fire Safety Design Methods, June 15–17, Lund, Sweden, 2000, pp. 93-103.

[17] T. Rinne, J. Hietaniemi, Use of Fire Safety Engineering and Fire Risk Analysis in Heritage Buildings, Proceedings of Interflam 2007, International Interflam Conference, September 3–5, London, England, 2007, pp. 1575–1580.

[18] Y. He, S. Grubits, A risk-based equivalence approach to fire resistance design for buildings, J. Fire Prot. Eng. 20 (1) (2010) 5–26.

[19] X. Huang, X. Wang, X. Jin, G. Liao, J. Qin, Fire protection of heritage structures: use of a portable water mist system Under high-altitude conditions, J. Fire Sci. 25 (3) (2007) 217–239.

[20] M. Green, Fire Safety Review of Heritage Buildings in NSW (Master Project Thesis) School of Computing, Engineering and Mathematics, Western Sydney University, 2009.

[21] C.A. Gorini, Master Math: probability, CourseTechnology/Cengage Learning, Boston, Mass, 2012.

[22] W. Mendenhall, R.J. Beaver, B.M. Beaver, Introduction to Probability and Statistics, 14th ed., Cengage Learning, Inc, Belmont, 2013.

[23] R.W. Hamming, Error detecting and error correcting codes, Bell Syst. Tech. J. 29 (2) (1950) 147–160 (MR 0035935).

[24] P. Jaccard, The distribution of the flora in the alpine zone, New Phytol. 11 (2) (1912) 37–50.

[25] E. Davenport, N. El-Sanhury, Phi/Phimax: review and Synthesis, Educ. Psychol. Meas. 51 (1991) 821–828.

[26] T.F. Cox, M.A.A. Cox, Multidimensional Scaling, 2nd ed., Chapman and Hall, London, 2001.

[27] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning, Springer-Verlag, New York, 2009.

[28] H. Akaike, A new look at the statistical model identification, IEEE Trans. Autom. Control 19 (6) (1974) 716–723. http://dx.doi.org/10.1109/TAC.1974.1100705 (MR 0423716).

[29] I. Kosmidis, D. Firth, Bias reduction in exponential family nonlinear models, Biometrika 96 (2009) 793–804.

[30] R. Mundry, Issues in information theory-based statistical inference - a commentary from a frequentist's perspective, Behav. Ecol. Sociobiol. 65 (2011) 57–68.