



**Escola de Camins**

Escola Tècnica Superior d'Enginyeria de Camins, Canals i Ports  
UPC BARCELONATECH

## MASTER THESIS

### Master

MEng Civil Engineering

### Title

Relating immission alarms with wind patterns  
in La Oroya valleys, Peru

### Author

Ryan Doolan

### Tutor

Agustí Pérez-Foguet

### Speciality

Environmental Engineering

### Date

17-May-2016

# Relating immission alarms with wind patterns in La Oroya valleys, Peru

---

Master Thesis

Ryan Doolan

17/05/2016

## Acknowledgement

This study would not have been possible without the support of various people. I must thank my thesis advisor Agustí Pérez-Foguet for his invaluable advice, guidance, and optimistic attitude throughout. I must thank my family and friends who endured this long process with me, from a safe distance, always providing support. They were there when I needed it, and stayed well enough away when I needed it more.

## Abstract

In this work, wind patterns in mountain valleys with complex terrain are compared with air pollution immission levels. The study presents a methodology for developing and testing a model to simulate SO<sub>2</sub> pollutant immission alarms, from information collected via monitoring stations in key areas around the town of La Oroya, based on the behaviour of local winds, categorised using previous work in this subject area and location.

The purpose of this research was to investigate immisions around the Doe Run metallurgical complex in La Oroya and to relate “emergency level” warnings with wind patterns. This will provide a foundation to a greater strategy to mitigate the risk associated with high emissions in this area, improving the health of the local residents without compromising their availability of work and potential economic prosperity.

The data used in this study was sourced from nine weather monitoring stations located around La Oroya and over a two year data collection period, from June 2007 to October 2008. Wind behaviour has been categorised into wind patterns based on work done in previous studies in this field and in this location.

This study has shown that the modelling of wind patterns and relating these to local immission alarm level can be done, however more data is required to update the model before it can have any operational use. Therefore the recommendations for further study in this topic firstly involve the collection of more data at the stations, as the more data collected means there is more chance of getting consecutive days of hourly data and will improve the accuracy of the model, and lessen the effects of monthly skew.

Finally, the findings presented in this study highlight the situation in La Oroya, Peru, where the main industrial complex provides both a large source of employment but also a large source of atmospheric emissions. Further developing the model presented in this work would allow operational changes to be made to the day to day emissions output of the complex, based on daily wind pattern analyses and immission alarm forecasts, and could be used as part of a wider strategy to ensure the residents of the town no longer having to make the choice between having economic prosperity or a good quality of life.

## Table of Contents

1.	Introduction.....	4
1.1.	Background .....	4
1.2.	Aims & Objectives .....	5
1.3.	Scope & Constraints .....	6
2.	Literature Review.....	7
2.1.	Definition of Air pollution .....	7
2.2.	Description of La Oroya & the complex .....	8
2.3.	Description of health deterioration in the town .....	9
2.4.	Work vs health.....	9
2.5.	Wind behaviours in mountain valleys.....	10
2.6.	Wind dispersion & thermal inversion .....	10
2.7.	Wind Pattern Analysis.....	11
2.8.	Wind patterns and Immissions .....	11
3.	Methodology .....	12
3.1.	Data collection and monitoring stations.....	12
3.2.	Wind Pattern Classification.....	14
3.3.	Using BN Modelling .....	15
4.	Results & Discussion .....	17
4.1.	Qualitative Results & description.....	17
4.2.	Results: Bayesian Network Modelling.....	26
4.2.1.	Explanation of data (alarm and wind pattern categories).....	26
4.2.2.	Testing the model: One station.....	27
4.2.3.	Testing the model: More stations .....	36
4.2.4.	Testing the model: All days .....	38
4.2.5.	Testing: More alarm options.....	39
4.2.6.	Final Model .....	40
4.3.	Limitations of Data.....	44
5.	Potential Application of the Model.....	45
6.	Conclusion & Recommendations .....	46
7.	References .....	47

# 1. Introduction

## 1.1. Background

The term “air quality” refers to the condition of the air around us. “Good” air quality means clear and clean air, free of pollution. Good air quality is vital to maintaining a healthy environment for water and soil, flora and fauna, and for human health. “Bad” air quality refers to air that has been polluted in some way, for example, emissions from natural events and from man-made sources. When a pollutant (or pollutants) reaches a large enough presence in the air (denoted by its concentration) to cause detriment to the environment and to human health, it is referred to as having poor air quality.

The average person inhales roughly 14,000 litres of air each day. Thus, the presence of a pollutant or other contaminant can have serious effects if consistently present in the air a person breathes. This is especially true of people already afflicted with respiratory issues, pre-existing medical conditions (particularly cardiac issues), as well pregnant women, the young, and the elderly.

In addition to human health, poor air quality can have negative repercussions on the environment. When pollutants come into contact with plants or animals, it can cause serious damage to the local ecology. When pollutants are dispersed and scattered from turbulent winds, or when heavier pollutants are settled out from calmer winds, they can land in water bodies or onto soil, which, in turn, can then be washed into waterways via rain and travel further. Poor air quality and pollutant presence can also have wider reaching effects on the climate with regards to warming, and is a large factor in the climate change narrative (European Environment Agency, 2013).

If there is a pollution related detriment to human health or to the environment, it can have negative economic impacts. Large costs can be incurred for hospitalisation/medical treatment, lost work days, and premature deaths. The damage to the environment can have less direct, but wider reaching effects, for example, damage to waterways, vegetation and soils can reduce the productivity of agricultural industries.

## 1.2. Aims & Objectives

The purpose of this research was to investigate immisions around the Doe Run metallurgical complex in La Oroya and to relate “emergency level” warnings with wind patterns.

Different statistical methods will be discussed and the most suitable selected in order to propose that the wind patterns can be used to provide a daily indicator of immission alarms.

This will provide a foundation to a greater strategy to mitigate the risk associated with high emissions in this area, improving the health of the local residents without compromising their availability of work and potential economic prosperity.

These aims will be fulfilled by completing the following objectives:

1. Record and track severity of emissions from the Doe Run metallurgical complex using immissions data from seven monitoring stations.
2. Collate hourly data from each station to identify any trends or “peak periods” for the exceedance of the “emergency level” limits. Categorise these occurrences.
3. Use wind pattern analyses from previous studies and compare them with days of immission limit exceedance, using statistical methods, to determine any relationship between wind patterns and alarm sounding.
  - a. Discuss statistical methods to relate wind patterns and immission alarms.
  - b. Develop model, based on statistical methods, to relate wind patterns and immissions alarms and explain findings.
  - c. Discuss potential applications of the model for the Complex with regards to forecasting alarms sounding and the use in some operative scheme for complex emission management.

### 1.3. Scope & Constraints

The data used in this study was sourced from the nine weather monitoring stations located around La Oroya. The stations were renovated in 2007 (D.E & B.O.C, 2007). This study uses hourly data for the period from July 2007 to September 2008, inclusive.

Additionally, there are many days within this two year period where data is collected non-consecutively (e.g. data is collected for 6-8am, omits 9am, and continues again at 10am) for whatever reason, precipitation can have an effect on the monitoring stations' effectiveness, for example. This then means that some days with missing hours do not give the entire picture of that day. Therefore, only days that have consecutive hours of data collection have been included.

Based on the availability of data, frequency of alarms and straightforward nature of daytime winds, the methods explored to correlate wind patterns and emission alarms will be done so for the morning hours initially, with potential to expand if the model is deemed suitable and effective.

The method will be developed using data for SO<sub>2</sub> presence only, with PM<sub>10</sub> being a potential avenue to explore at a later time, when testing the effectiveness of the model or simply expanding it.



## 2. Literature Review

### 2.1. Definition of Air pollution

Air pollution is the presence (or introduction) of a substance or particle into the air which has harmful effects on the local environment. It is one of the most prominent issues currently facing societies across the world, especially in areas situated near large roads and/or urban centres (vehicular emissions) and industrial centres (industrial emissions). Air pollution causes degradation of human health, damages the natural and built environment, and is also a contributing factor to the wider issue of climate change.

Main pollutants:

- **Sulphur Dioxide (SO<sub>2</sub>)** is a gas which reacts easily with other substances and can form harmful compounds, such as sulphuric acid, sulphurous acid and sulphate particles.

The extreme majority (99%) of the SO<sub>2</sub> in the air originates from human sources. The main such source is industrial activity, especially activity involving the processing of materials that contain sulphur. Certain mineral ores contain sulphur, making them one such example of a source of SO<sub>2</sub>, released during their processing. Additionally, industrial activity that involves the use or combustion of fossil fuels can also be a significant source of SO<sub>2</sub>. Sulphur dioxide is also present as a result of vehicular emissions, due to the combustion of fuel in the engine. (Australian Gov't, 2005).

- **Particulate Matter (PM)** is a mixture of extremely small particles and liquid droplets and consists of various components, such as metals, organic chemicals, acids (e.g. nitrates, sulphates etc), and soil/dust particles.

The size of these particles is directly proportional to the potential they hold to cause problems to health. Of particular concern are those particles of diameter 10 micrometers or less (PM<sub>10</sub>) as those are small enough to enter the body through the throat and/or nose and settle in the lungs. Once inhaled, these particles can affect the lungs and heart and have serious repercussions on human health (EPA, 2010).

The U.S. Environmental Protection Agency (EPA) groups particle pollution into two categories:

- **Inhalable coarse particles:** found near roadways and dusty industries, are larger than 2.5 micrometers and smaller than 10 micrometers in diameter.
- **Fine particles:** found in smoke and haze, are 2.5 micrometers in diameter or less. These particles can be direct results of emissions from sources such as forest fires, or when gases emitted from power plants, industries and automobiles react in the air.

Additionally, some pollutants undergo changes when they mix with other sources of air pollution. These chemical reactions result air pollutants referred to as ‘secondary pollutants’ instead of the ‘primary pollutants’ that are emitted from the point source and into the air.

The World Health Organisation (WHO) and the International Agency for Research on Cancer (IARC) denote airborne particulates as a Group 1 carcinogen. In 2013, a study involving over three hundred thousand people across nine countries in Europe revealed that there is no level of particulates that is considered “safe” and for every increase in  $10\mu\text{g}/\text{m}^3$  of PM10 present in the ambient air, the lung cancer rate rose by 22%. The instances involving lesser diameter particles of PM2.5 had a 36% increase in lung cancer for the same level of increase, highlighting the danger here as it can penetrate more deeply into the lungs (Raaschou-Nielsen et al, 2013).

## 2.2. Description of La Oroya & the complex

La Oroya is a smelter town, located in the central the Andes in the heart of Peru. The area of La Oroya is located in the province of Yauli and part of the territory of Junín, an area known historically for its agricultural output and abundant ore and mineral resources. It is at 3,740 metres above sea level and has an area of  $388\text{km}^2$ .

The La Oroya metallurgical complex was built in 1922, where it was initially only a copper smelter, then in 1928 a lead smelter was added and then a zinc refinery was built in 1952 Annual capacity of copper was 70,000 tonnes; lead was 122,000 tonnes; and zinc was 45,000 tonnes. However, production could be kept below these levels to remain below emission limits or to avoid the temperature inversion effect trapping gases over the town and local area (U.S. S.E.C., 2006).

Local mining operations nearby produce “dirty concentrates” ie their output contains impurities that the normally flotation process cannot separate or filter out. Thus, over time the La Oroya metallurgists have been able to use different methods to separate out these metals and recover them as a byproduct. The main smelters have been heavily integrated to this process and as a result, La Oroya produces gold and silver (from refinery residues) antimony, arsenic trioxide, bismuth, cadmium, indium, selenium, tellurium, sulfuric acid and oleum (U.S. S.E.C., 2001).

As a result of this technology, the smelters’ overall emission of certain noxious and toxic metals has reduced; however, as it has made the plant more complexly integrated, it is more difficult to modify individual aspects of the operation.

Doe Run acquired the complex in 1997 and with it, a “complicated and semi-obsolete smelter complex” according to the Guardian (2007). “The operation had suffered from disrepair and previous owners had invested little in modernisation or clean operations. As a result of years of pollution, the hills immediately around the

smelter became completely denuded, the river became more toxic, and the health of area inhabitants suffered”.

### **2.3. Description of health deterioration in the town**

A study in 1999, by the Peruvian Ministry of Health Environmental Health Directorate (DIGESA), released the results of a study on blood lead levels (BLL) in La Oroya. This study found that 99.1% of children under 10 years of age that were tested had BLLs higher than 10ug/dl, which is the maximum permissible level established by the World Health Organization (WHO). Childhood lead poisoning has extremely adverse effects including neurological system damage, poor intellectual performance, gastrointestinal and respiratory diseases, cancer and early death (Bellinger and Bellinger, 2006; Lidsky and Schneider, 2006). The study confirmed the smelter to be the prominent source of the pollution. It was also confirmed that, unless the source of the pollution was effectively controlled, or the affected residents moved from the pollution area, medical treatment would continue to be ineffective.

Another study, by the Consortium for Sustainable Development in La Oroya, returned similar results with regards to blood lead levels in children, and also reported high levels in pregnant women. When tested, children returned an average BLL of 39.49ug/dl, and pregnant women returned an average of 41.81ug/dl (UNES, 2000). A year-long study from June 2004 to 2005 also determined that newborn babies in La Oroya had an average of 8.84ug/dl BLL (Pebe, et al., 2008).

In addition to air pollution, evidence of contamination has also been found inside people’s homes, bringing to the fore the case of a long-term housing problem in La Oroya (Cornejo and Gottesfeld, 2004).

### **2.4. Work vs health**

In the past decade, La Oroya town has been subject to significant international attention and scrutiny due to reports of an extremely high number of children having high levels of lead poisoning as a direct result of pollution from the smelting complex. However, instead of the local populace appealing to the government or using the international attention garnered to protect the health of themselves, their children, and the environment, the majority of the community chose to minimise the issue with a view to defending the smelter and retaining job prospects there.

Due to this paradoxical situation, this catch-22 between health and economic prosperity, the La Oroya case has become known as a frontrunner in the health vs work argument in the problem of human rights (Scurrah et al, 2009; Valencia, 2012).

## 2.5. Wind behaviours in mountain valleys

Valleys in mountainous environments are extremely interesting areas with regards to airflow and wind behaviour, due to the complex nature of the terrain and the nature of the 'forcing mechanism' (either mechanically or thermally) causing the air to move (Drobinski et al, 2003).

Towns and cities in areas of limited ventilation (valleys, mountain valleys etc) can experience severe air pollution and therefore concern for public health (Romero et al, 1999; Edgerton, 1999; Panday and Prinn, 2009). The transportation of the particles causing this air pollution out of the valley atmosphere can be dependent not only on the orography and barriers associated, but can also be limited by air circulations by the urban heat island effect (Haeger-Eugensson and Holmer 1999; Savijärvi and Liya 2001), or the presence of near-surface level temperature inversions (Janhall et al, 2006; Yao and Zhong, 2009).

Locally, urban land characteristics can have a significant influence on the heat transfer between the land surface and the atmosphere (Arnfield, 2003; Grimmond, 2007; Kanda, 2007). Thus, changes in land use or land cover as a result of urbanisation can modify the dynamics of local wind and air flow (Lee and Kim, 2008), the urban heat island (Chen et al, 2006), and temperature inversions (at a local scale) which can impact air quality and have repercussions for future urban planning works (Romero et al, 1999; Janhall et al, 2006; Rizwan et al, 2008; Allen et al, 2011).

## 2.6. Wind dispersion & thermal inversion

When pollutants are emitted and released into the local atmosphere, the main factor in determining how well they disperse is the behaviour of the surrounding air, ie wind flow. In this way, turbulent flows will quickly mix pollutants into the air. For example, during warmer days (summer) the air near the surface can have a higher temperature than the air above, causing it to rise. If large amounts of the warmer air rise up, rapid mixing can occur.

The speed of the wind is also a contributing factor to how quickly and how far a pollutant can be transported away from its source, though strong winds cannot always completely disperse a pollutant.

In contrast, sometimes atmospheric conditions are very stable and so mixing does not occur. This can happen if, contrary to the first example, the air at the surface is colder than the air above (known as temperature inversion). The colder air does not rise and therefore mixing does not occur. If a pollutant is released at surface level it will be trapped under the 'cool layer' of air and will begin to build up. Temperature inversions can persist throughout the day in winter (BCME, 2015).

## 2.7. Wind Pattern Analysis

Identifying and analysing the wind patterns that characterise air flow in a valley is a way to understand the nature and configuration of the wind field in a set area, especially one in complex or closed-off terrain, and is extremely important for studies pertaining to air quality and pollutant transport (Darby, 2005; Ratto *et al.*,2010).

In a previous study for La Oroya, Perez-Foguet (2014) used cluster analysis (CA) on hourly wind data collected from nine monitoring stations located around the town and valleys. This was then used to characterise the wind patterns through separate ‘clustering’ of wind speed and wind direction. By separating them, the influence each characteristic (ie speed or direction) has on the data returned can be identified independently, giving parameters to identify/classify wind patterns using cross-tabulations between heat maps and multiple contingency tables.

## 2.8. Wind patterns and Immissions

Information gathered via wind pattern analyses can be used in immissions modelling studies also as, if the behaviour of the wind that transports the pollutants can be analysed, then the presence of pollutants and extent of transport/dispersion can be forecasted. A study by Arroyo *et al* (2009) took an approach using wind patterns, analysed and categorised using similar criteria as above (i.e. wind direction, wind speed, dry temperature, relative humidity, atmospheric pressure and solar radiation) then went on to use these wind patterns to model immision values in urban environments.



### 3. Methodology

An overview of the town of La Oroya (location, population, main industries) and its situation with regards to air quality and pollution was drawn up, along with an explanation of the pollutants that are present in the air and are to be studied in this report, SO<sub>2</sub> and PM<sub>10</sub>, and their effects on human health and the environment. The argument of “Work vs health” was also introduced.

#### 3.1. Data collection and monitoring stations

For primary information of air pollutant presence, data was taken from seven monitoring stations set up at various locations in the valleys of La Oroya town. These locations are as follows:

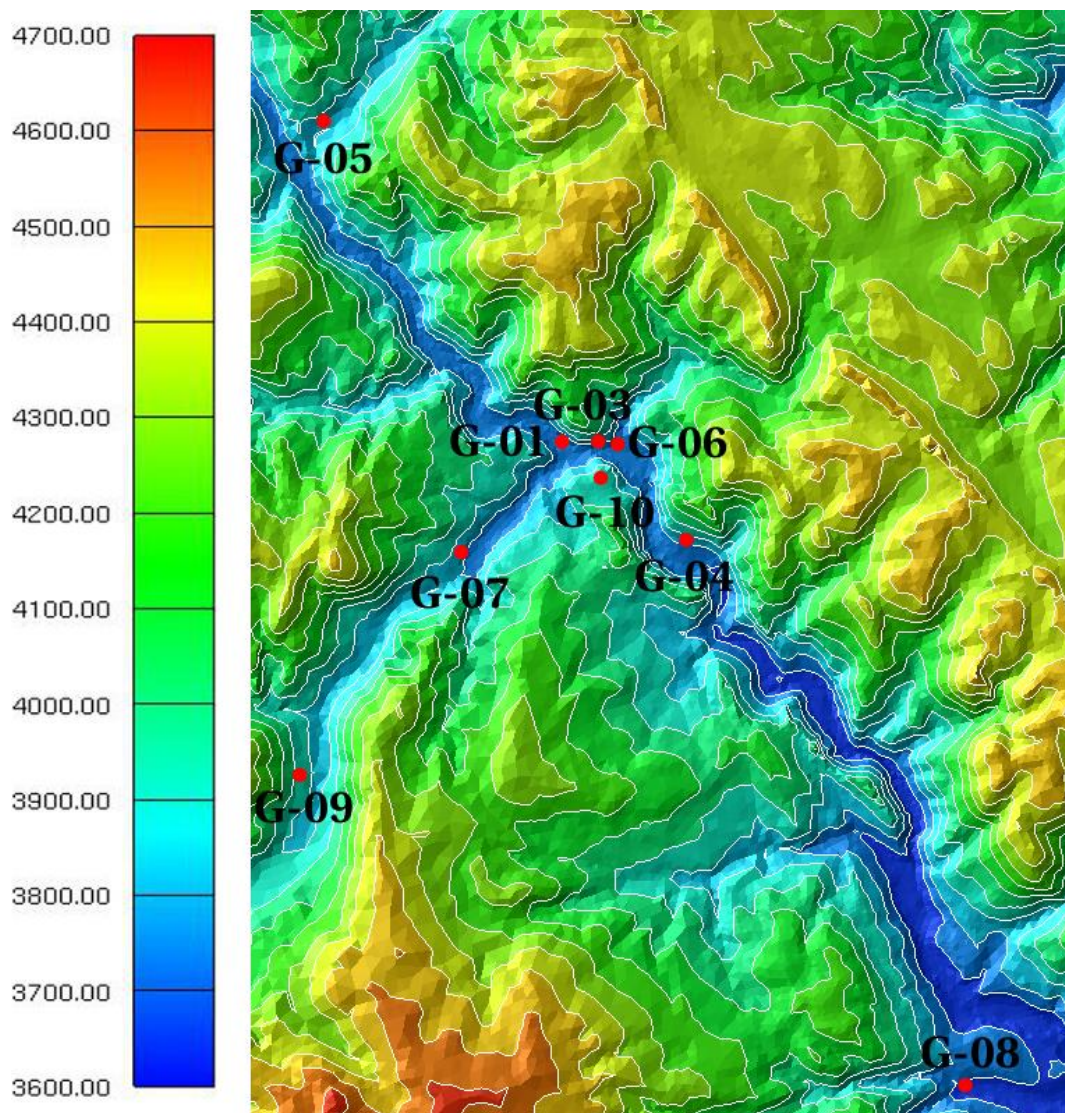


Figure 3.1 - Orographical map of La Oroya with monitoring stations superimposed (source: Perez-Foguet, 2014)

Table 3.1 - Summary of monitoring station information (source: D.E. & B.O.C., 2007; C.N.A, 2007)

Label	Name	Altitude	Location
G01	Hotel Inca	3734 m	Central Zone
G03	Sindicato de Obreros	3736 m	Central Zone
G04	Huanchan	3790 m	SE Valley
G05	Casaracra	3779 m	NW Valley
G06	Fundicion	3721 m	Central Zone
G07	Marcavalle	3780 m	SW Valley
G08	Huari	3684 m	SE Valley
G09	Huaynacancha	3860 m	SW Valley
G10	Cerro Sumi	3936 m	Central Zone

The locations of the nine stations are denoted in Figure 3.1 (previous page) and their details summarised in Table 3.1. Station G10 is elevated; located on the rise above the main intersection of the two valleys. Stations G01, G03, and G06 are located at the bottom of the central valley, along the main intersection between the two valleys. The rest of the stations are located at various points along the valleys, with G05 being at a point far NW; G08 being at a point far SE; G09 SW; and G04 and G07a shorter distance from the main intersection SE and SW respectively.

The intended usage for these stations, and therefore available data, are as follows: G06 and G10 are weather monitoring stations; G01, G03, G04, G05, G07, G08, G09 are the seven stations recording local pollutant presence. The data for wind direction, speed, and local temperature were available from every station and, additionally, data pertaining to solar radiation and atmospheric pressure were available from station G06.

All of the hourly data for pollutant content (SO<sub>2</sub>) for each station collected during the years 2007 and 2008 was collated and statistically processed to give the mean values and variance. This was done for the “Total data”, ie all of the data collected over the two years and inclusive of non-consecutive hours, in order to generate a graph of SO<sub>2</sub> levels per hour.

The “Total” graph was used, with consecutive and non-consecutive hours alike, to increase the number of data points available and to make trendlines in the graph more apparent. It also smooths over any spikes in data due to seasonal changes etc to give an idea of how a ‘typical’ day looks.

The same was then done for long collection periods of consecutive hours, ie data collected every hour – uninterrupted – for as long as possible, to give a more accurate view of certain days, ie it shows exactly what happened each hour of a

particular day, not just an average or the theoretical 'typical' day. This was carried out for July '07 and for March '08, to highlight seasonal changes between winter and summer, respectively.

The graphs were analysed qualitatively with general trends identified and possible reasons for these trends, and for peak hours etc, being speculated on in the descriptors. Seasonal differences were highlighted and commented on also, as were the general relationships between stations (location, distance, altitude etc) and qualitative descriptions of wind behaviour and their potential for carrying the pollutants to other locations, mainly other stations.

### 3.2. Wind Pattern Classification

Perez-Foguet's study (2014) in categorising wind patterns in complex mountain valleys, using the same weather stations as this report, was used to supply the wind pattern 'signature' (11; 12; 52; etc) based cluster analyses of wind speed and wind direction. The hourly data was then refined to just these signatures such that, assuming the initial data was available, each hour also included a wind pattern (WP) or category. It is these wind patterns that will be used in the comparison between high immisions in stations (ie alarms) and wind behaviour (ie wind patterns).

For the in-depth quantitative analysis, three methods were explored to check for any kind of linear relationship between wind patterns and high immisions readings, in this case categorised by an alarm level (No alarm; Level 1; Level 2; Level 3) which is, in turn, dictated by government-set limits (Limit 1 =  $500\mu\text{g}/\text{m}^3$ ; Level 2 =  $1500\mu\text{g}/\text{m}^3$ ; Level 3 =  $2500\mu\text{g}/\text{m}^3$ ). If the emission level, for example  $\text{SO}_2$  content, is above one of these limits for three hours an alarm will sound.

These three methods were:

- **Multinomial Logistic Regression**

Multinomial Logistic Regression is used when the dependent variable is nominal, in this case categorical and with more than two categories, and the set of categories are not ordered in any meaningful way.

This method was deemed to be unsuitable as, for ordinal categorical variables; the drawback of the multinomial regression model is that the ordering of the categories is ignored.

- **Ordinal Logistic Regression**

The ordinal regression model is very similar in nature to the multinomial model introduced previously, however the categories are ordered in a meaningful way, with the order of importance set by the user. So it can be simply "lowest to highest" but can also be in some special order that the user has defined.



However, the drawback of the ordinal regression model is that it isn't well adjusted to many variables (ie many categories) which makes it unsuited for the requirements of the study, as the input has six variables with between six and ten levels. This makes the model too complex and unwieldy to apply quickly or easily.

- **Bayesian Network (Modelling)**

A Bayesian network model is a statistical model that represents a set of random variables and their conditional dependencies via probabilities and a graphical output. It is used to find relationships between data that are categorical in nature and can range from

It separates each data series or category into nodes, and then assigns connecting lines between nodes to show some kind of correlation, with the direction of the arrow showing some kind of causation. Standalone nodes with no lines have no influence on other nodes.

### 3.3. Using BN Modelling

There were many options for the nature of Bayesian Network model that checked for different things in different ways. A combination of options were that are useful and suited to categorical data were selected, with these options being explored to investigate which option was more comprehensive and the best fit to the data and for the requirements of the study.

For example, using different functions within the "BNlearn" package in the R program (R Core Team, 2016; Scutari & Denis, 2014) used to investigate these connections in the data will check for different indicators of connection. This is shown in the comparison of "score.bde" and "score.k2" below:

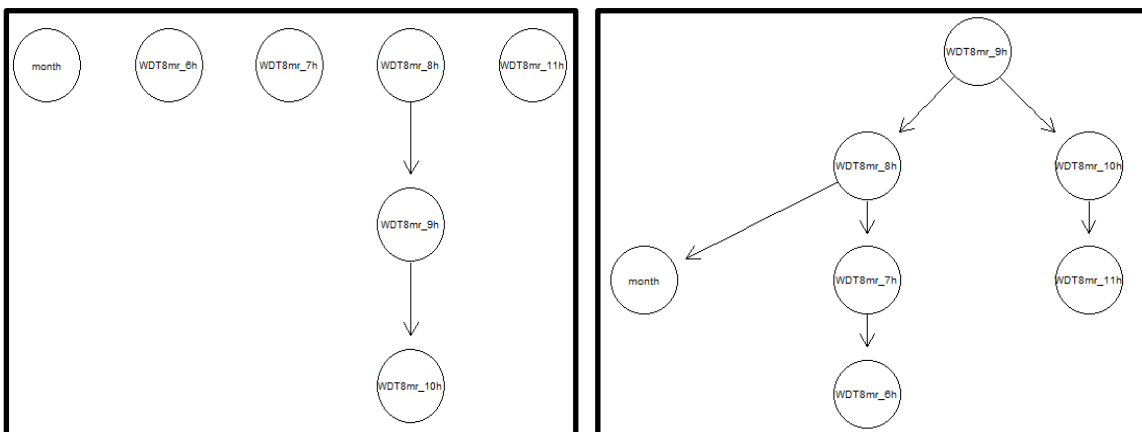


Figure 3.2 - Comparison between "bde" (left) and "k2" (right) methods

From Figure 3.2, above, it can be seen that using the “bde” option only identifies a connection between the wind patterns between 8am and 9am; and 9am and 10am. It does not identify any connection between any other categories. Whereas option “k2” at least some kind of connection between each node, and though some of the connection flow the “wrong” way (e.g. the wind at 8am influencing the wind at 7am, which of course does not make sense logically) it does recognise a connection. It is then up to the user to apply logical or expert knowledge to the model to identify the correct nature of the connection. This basic example of how the model works serves to explain why the k2 technique was used throughout.

After the modelling method was decided, a tester for the model was created. Using an initial period of eleven days in July and only using one station (G01), for simplicity, the relationship between wind patterns hour to hour (from morning winds) and between hourly wind patterns to alarm levels was investigated. At this stage, because of the small number data points, the probabilities of the different outcomes can be shown and discussed also.

The following sections show how the model grew with the addition of the other stations, more days (first for July only, and then for a few days of each month), before reaching a point where it can be considered useful. The probabilities are not included for each of these stages, however, as, though they are necessary to explain the model, they are only really relevant at the first stage (to explain the model when it is still simple with relatively few data points) and at the final stage, when it can be discussed what to do with the predictions in the model to improve the situation at La Oroya.

After this, more days were checked for and with more stations. The stations chosen typically had alarm behaviour during the midday hours only thus were deemed to be similar in nature. These results can then be checked for (a) similarities to the initial model, i.e. which relationships stay constant; (b) differences to the initial model, i.e. which relationships change when more data is available; and (c) new relationships due to the addition of other stations and how their inclusion affects the model.

The model was then checked for all available days for all stations, excepting station G04 as it very often had an alarm and was mostly at Level 3. This high alarm content meant that model returned many connections to this station, and other connections to alarms were deemed to be ‘weaker’ and were ‘wrapped up’ in, or eclipsed by, the G04 connection. The model was then checked for fewer days, but for a few months with more variability in alarm presence and level. This was done to determine the WP to alarm structure.

Finally, a combination structure using the previous two (all days for winds, three months for alarms) was developed. This was deemed to present the most accurate account of the relationships between nodes based on the two years of data available.

The two aims of the iterative modelling approach, i.e. improving the model structure each time by adding more data and applying background knowledge; were to determine the relationship between hourly wind patterns (so that by knowing what happens at 6am, for example, there is a reasonable chance to predict the WP possibilities at 7am) and to determine the relationship between hourly WPs and alarm presence. Hence, more days were used for determining the wind pattern structure, and, as it turned out, fewer days were used to determine the alarm structure.

## **4. Results & Discussion**

### **4.1. Qualitative Results & description**

The pollutant content (SO<sub>2</sub>) data recorded by each station is shown via three graphs. One showing the total data for all of the collected data points over the two year period, regardless of continuity/ consecutive days; the second shows the hourly data for a nine day consecutive recording period in July 2007, to represent hourly activity during a typical winter month; and the third shows the hourly data for a ten day consecutive recording period in March 2008, to represent hourly activity during a typical summer month.

Data graphs are included for each station monitoring immissions, the government-imposed limits of SO<sub>2</sub> immission are superimposed onto the images, and the breakdown is as follows:

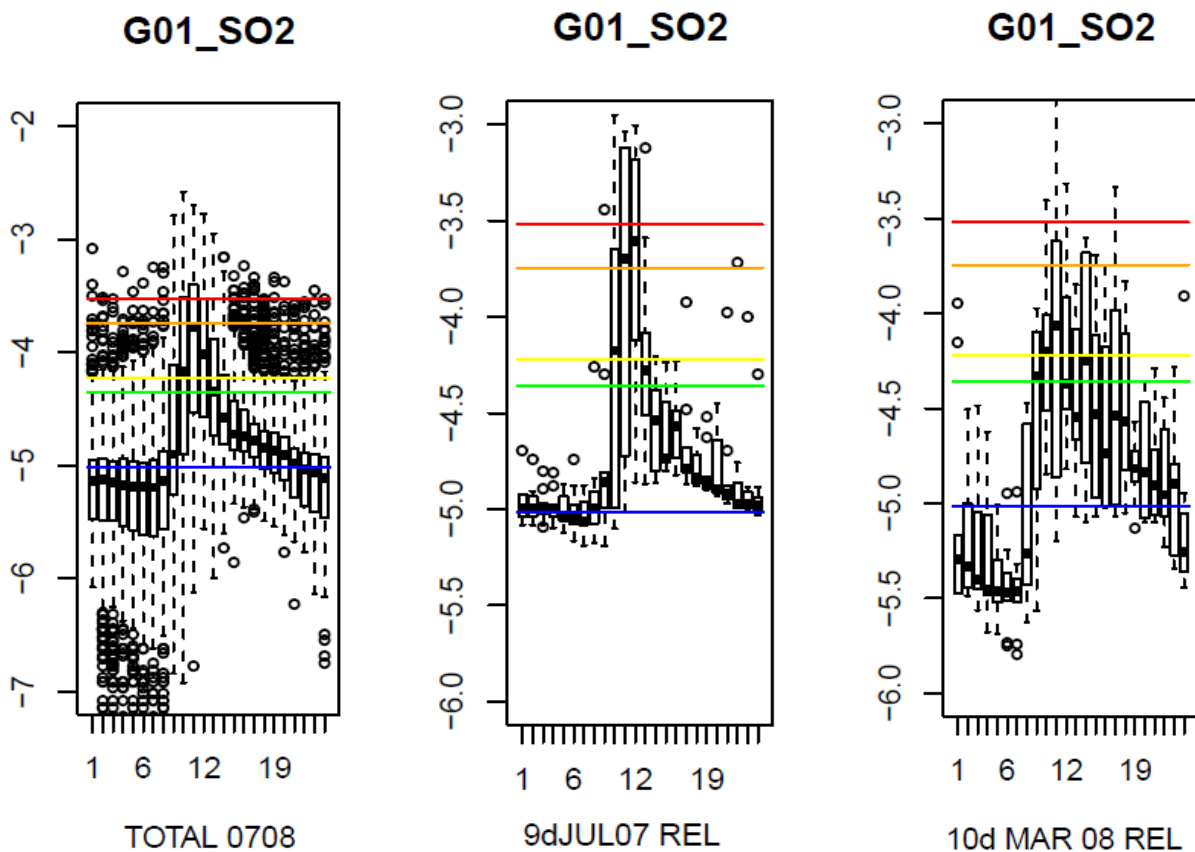


Figure 4.1 - Immission levels ( $1 \times 10^0$ ) vs time (hours) for Station G01 for: Total data (a); July 07 (b); March 08 (c)

Figure 4.1 (a), above, shows the hourly data collected by Station G01 during the entire monitoring period and shows that, during a typical day, the mean value of SO<sub>2</sub> content present in the air is between the first and second ‘emergency’ emission levels during the midday hours, before reducing again hourly from the late afternoon through the evening, before spiking again at midday.

Figure 4.1 (b), above, shows the hourly data collected during 9 consecutive days (days of uninterrupted data every hour with no omissions) and represents the winter months in terms of air flow. It is a very similar pattern to the first graph (with less variation and fewer outliers due to fewer data points), with the ambient SO<sub>2</sub> content in the morning a steady, low-level presence, before rising sharply at midday to level 2 “Emergency Level” SO<sub>2</sub> content, and then falling steadily through the afternoon to settle in the evening through the morning again.

Figure 4.1 (c) above, shows the data collected during 10 consecutive days of hourly immissions monitoring and represents the summer months for the air flow in the valley. It is similar in its general trend to the first two graphs, with the steady level of immissions in the morning and spiking in the afternoon (though spiking roughly an hour earlier), however the SO<sub>2</sub> in the air lingers for longer than the other two. It still

reduces steadily in the late afternoon, though at a lower rate, meaning there is more SO<sub>2</sub> present in the air in the early evening, though within emergency levels.

Station G03 – Sindicato de Obreros

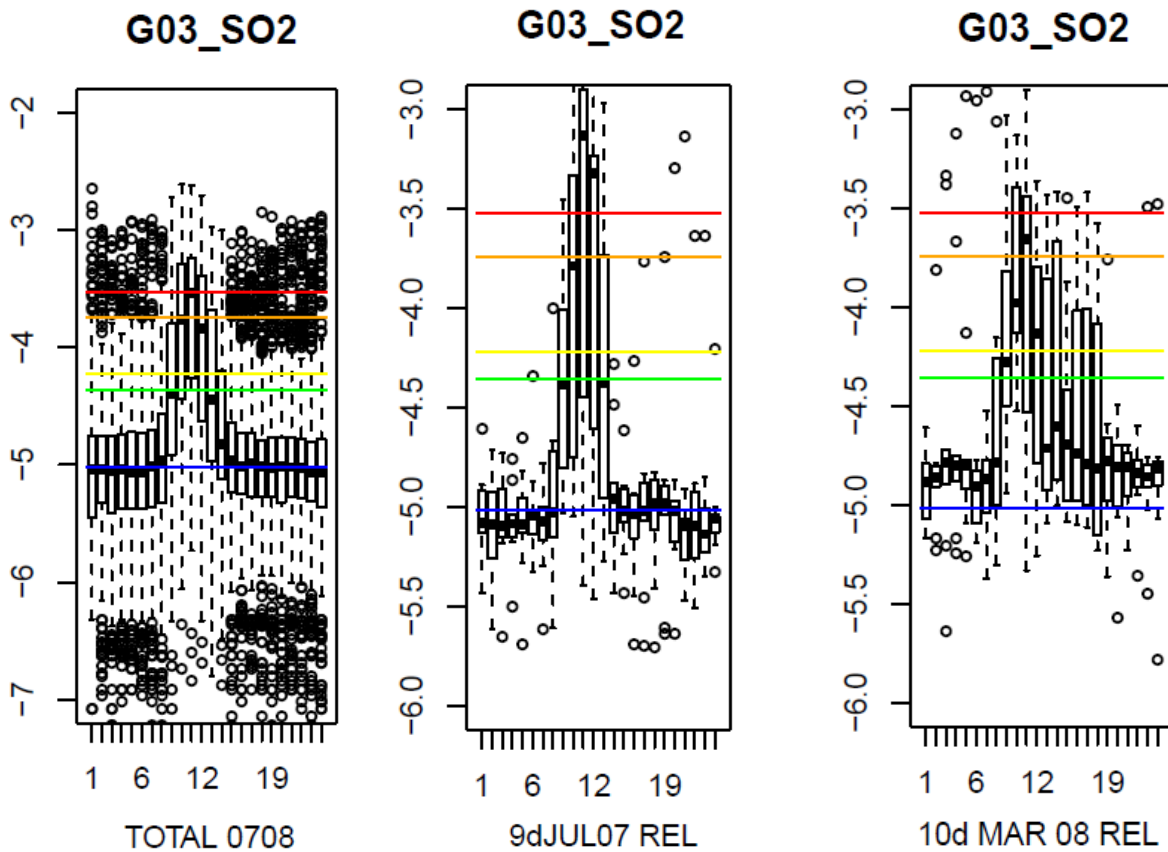


Figure 4.2 - Immission levels ( $1 \times 10^{11}$ ) vs time (hours) for Station G03 for: Total data (a); July 07 (b); March 08 (c)

Figure 4.2 (a) above shows the hourly data collected by Station G03 during the entire monitoring period of 2007 and 2008, with the darker (black) squares representing the mean value for that hour. It shows that, during a typical day, the late night/early morning SO<sub>2</sub> immissions are fairly constant and at a low level. After, 8am, the SO<sub>2</sub> level rises rather sharply, passing through the first alarm level and hovering near the second (at 10am), before reaching its peak at 11am hovering at the third (and highest) alarm level. It then falls to just below the second alarm level at 12pm; and falls over the next three hours to settle back into the low level, consistent SO<sub>2</sub> level at 3pm.

Figure 4.2 (b) above shows the hourly data collected by Station G03 during a nine-day continuous period (days of uninterrupted data every hour with no omissions) in July of 2007 and is a good indicator of the winter months and how the air flow (and thus, transport of pollutants) is affected by seasonal changes, in this case lower temperatures and precipitation.

The data collection of the stations themselves can also be affected (likely due to the effects of precipitation). This graph shows a very similar pattern to the first, (though with more abrupt changes between each hour, as there are far less data points to smooth out the trendline) with late night & early morning immissions being constant and low level. This then begins to rise rapidly after 8am (as before), triggers the Level 1 alarm at 10am (and is very close to Level 2), triggers the Level 3 alarm at 11am and sustains this for one hour, and then drops just as rapidly to a “No alarm” state at 1pm before returning to the low level from 2pm where it fluctuates slightly, but remains very low until 8am the next day.

Figure 4.2 (c) above, shows the data collected during 10 consecutive days of hourly immissions monitoring in March 2008 and represents the effect of the summer months, with increased temperatures and precipitation in the valleys, and the subsequent effect on air flow and pollution. It is similar in its general trend to the first two graphs, with the steady level of immissions in the morning and spiking in the afternoon, however the SO<sub>2</sub> in the air lingers for longer than the other two. It reduces steadily in the late afternoon at a lower rate, meaning there is more SO<sub>2</sub> present in the air in the early evening though still well within ‘safe’ levels. Also, though it is still at a very low order of magnitude, there is a higher level of ambient SO<sub>2</sub> in March than there is in July.

Station G04 – Huanchan

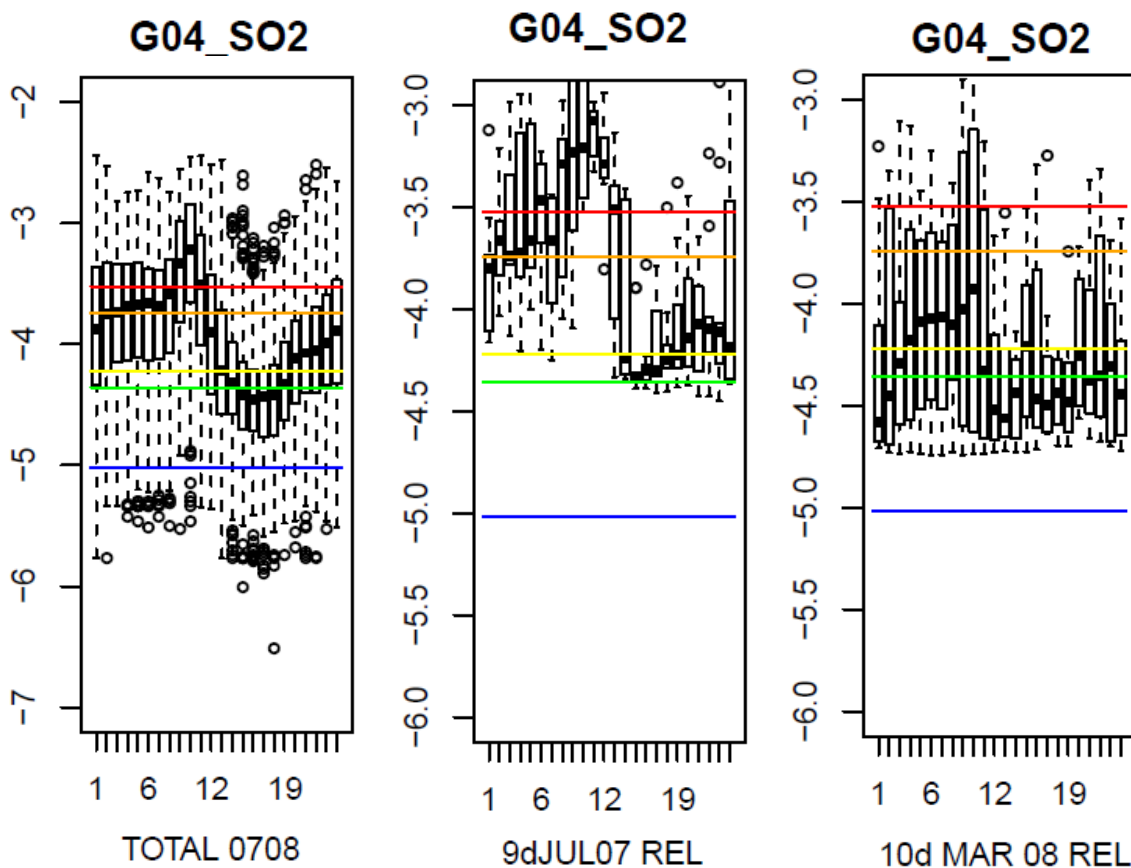


Figure 4.3 - Immission levels (1x10<sup>9</sup>) vs time (hours) for Station G04 for: Total data (a); July 07 (b); March 08 (c)

Figure 4.3 (a) above shows the hourly data collected by Station G04 during the entire monitoring period of 2007 and 2008, with the darker (black) squares representing the mean value for that hour. This graph shows an obvious trend with little variance between hours, but a steady increase or decrease with small changes hour to hour. It does, however, record levels much higher than that of the other stations, with only six hours with no alarm.

The lowest level of SO<sub>2</sub> content occurs at 4pm and steadily rises until alarm Level 1 is sounded at 8pm, where it continues to rise steadily throughout the night, reaching Level 2 at 2am, where it continues to rise steadily, before spiking slightly at 8am and then spiking again at 9am where it reaches alarm Level 3 and sustains this level for two hours (with the highest level reached at 10am). The level then falls significantly between 12pm and 2pm, and then falls steadily after that, with the alarm stopping at 3pm, before reaching the lowest point at 4pm again.

Figure 4.3 (b) above shows the hourly data collected by Station G03 during a nine-day continuous period (days of uninterrupted data every hour) in July of 2007 and is a good indicator of the winter months and how the air flow (and thus, transport of pollutants) is affected by seasonal changes. The data collected by the stations themselves can also be affected.

This graph shows an overall similar pattern to the first, though with far more abrupt changes between each hour, with a low level at late-afternoon/early evening and then rising throughout the night and early morning until reaching a peak at midday, before dropping again. However, the midday to early evening drop is much more extreme, and the early to late morning rise is much more pronounced, with the overall SO<sub>2</sub> content being higher.

Figure 4.3 (c) above, shows the data collected during 10 consecutive days of hourly immissions monitoring and represents the summer months for the air flow in the valley. The early morning to midday rise is a very similar trend to the July (winter) graph, though it begins at a much lower SO<sub>2</sub> content. The increases and decreases in SO<sub>2</sub> are more gradual and steady, and after the midday peak throughout the late afternoon and the evening, there is a small but noticeable fluctuation with levels rising and falling slightly until early morning where they begin to rise again for the midday peak. During this period (summer) there is a greater number of hours spent with no alarm sounding, and when it does, there are no Level 3 alerts.

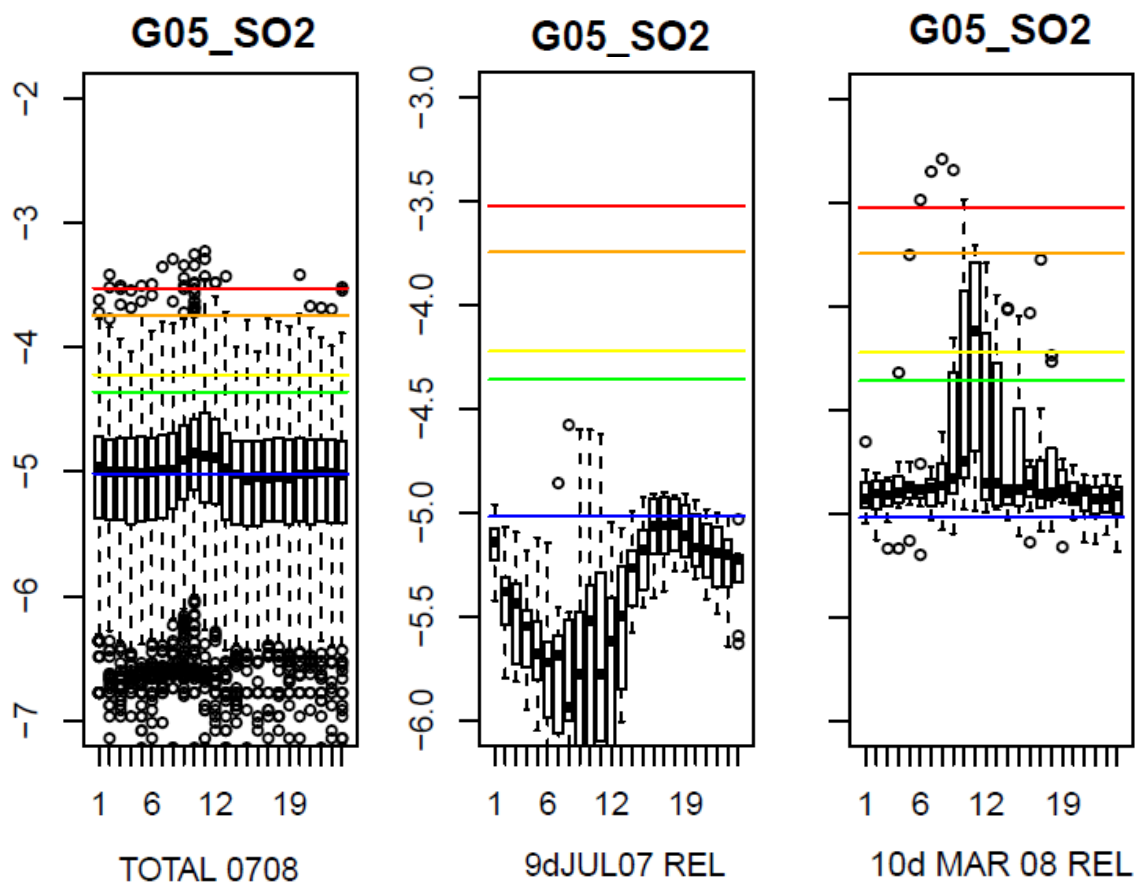


Figure 4.4 - Immission levels ( $1 \times 10^9$ ) vs time (hours) for Station G05 for: Total data (a); July 07 (b); March 08 (c)

Figure 4.4 (a) above shows the hourly data collected by Station G05 during the entire monitoring period of 2007 and 2008, with the darker (black) squares representing the mean value for that hour. This graph shows a horizontal trend with little variance between hours but for a slight increase over the midday hours, in keeping with the midday spike observed in the other graphs, however, it is still very low level and well under the alarm threshold.

Figure 4.4 (b) above shows the hourly data collected by Station G05 during a nine-day continuous period (days of uninterrupted data every hour) in July of 2007 and is a good indicator of the winter months and how the air flow (and thus, transport of pollutants) is affected by seasonal changes. The data collection of the stations themselves can also be affected (likely due to the effects of precipitation).

This graph shows a different pattern to the first, it does not hover around a constant level, but fluctuates during the day; falling throughout the late evening and early morning, rising from mid-morning (6am) with some inter-hour fluctuation, but still on a general rising trend until early evening (6pm) where it begins to fall again for the rest of the night.



Figure 4.4 (c) above, shows the data collected during 10 consecutive days of hourly immissions monitoring and represents the summer months for the air flow in the valley. It is similar in its general trend to the first graph, with the steady level of immissions in the morning and spiking in the afternoon, however the ambient SO2 levels in the morning and evening are slightly greater and the midday spike is certainly more pronounced (and in line with other stations' data for this time) and actually triggers the Level 1 alarm.

Station G07 – Marcavalle

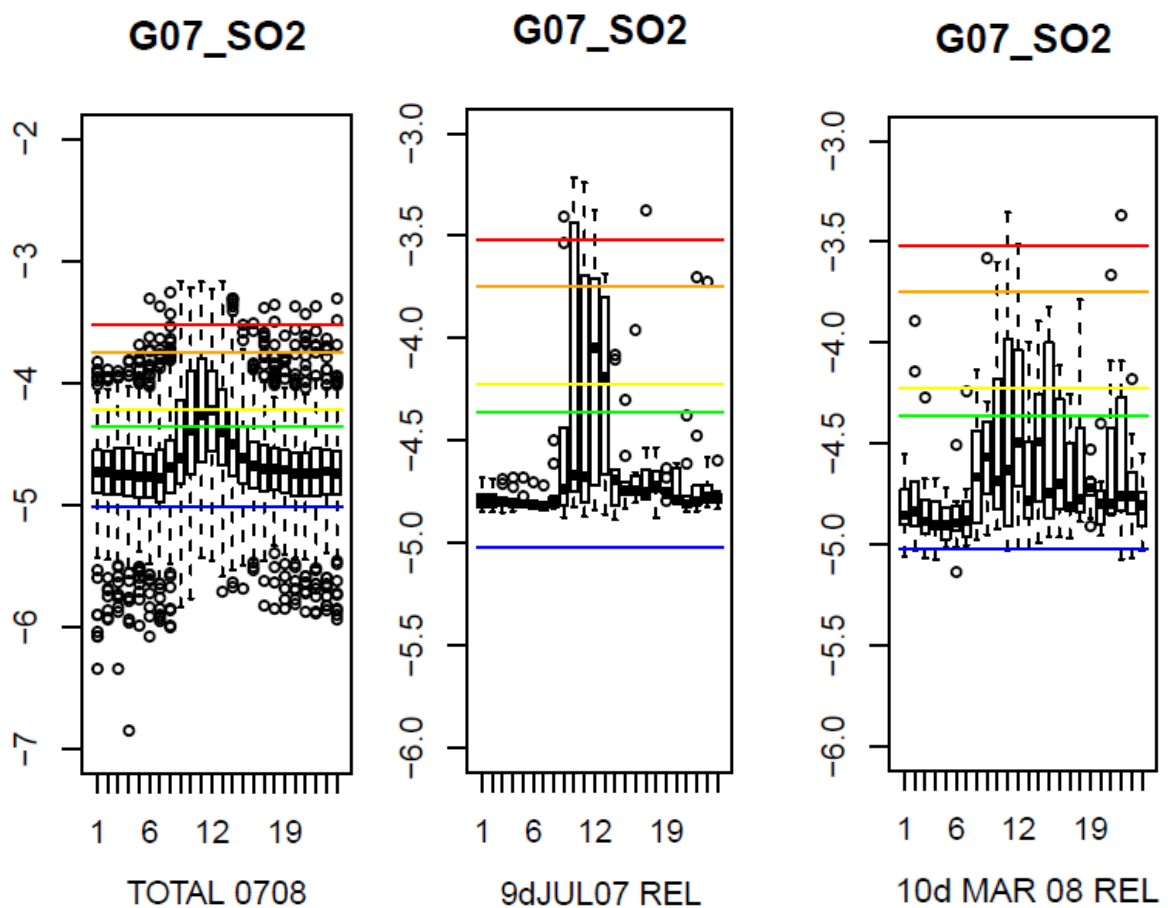


Figure 4.5 - Immission levels ( $1 \times 10^9$ ) vs time (hours) for Station G07 for: Total data (a); July 07 (b); March 08 (c)

Figure 4.5 (a) above shows the hourly data collected by Station G07 during the entire monitoring period of 2007 and 2008, with the darker (black) squares representing the mean value for that hour. This graph is similar to G05, G03, and G01 (though G02 is skewed slightly) in having the classic “bell curve” shape with fairly constant SO2 levels in the evening and morning hours and a rise in the midday, peaking at 12pm and being on the cusp of the Level 1 alarm.

Figure 4.5 (b) above shows a similar situation to the first, with late night & early morning SO2 levels fluctuating slightly but at low levels, but the midday spike is much more abrupt and rises very sharply between 11am and 12pm, where it trips

the Level 1 alarm, remains at that alarm level for one hour, before falling very sharply again at 2pm, where it fluctuates very slightly throughout the night.

Figure 4.5 (c) above, shows the data collected during 10 consecutive days of hourly immissions monitoring and represents the summer months for the air flow in the valley. It is similar to the first two graphs for the early morning hours only, with the rest of the day fluctuating constantly, though not being a high enough level to trip any alarms.

Station G08 – Huari

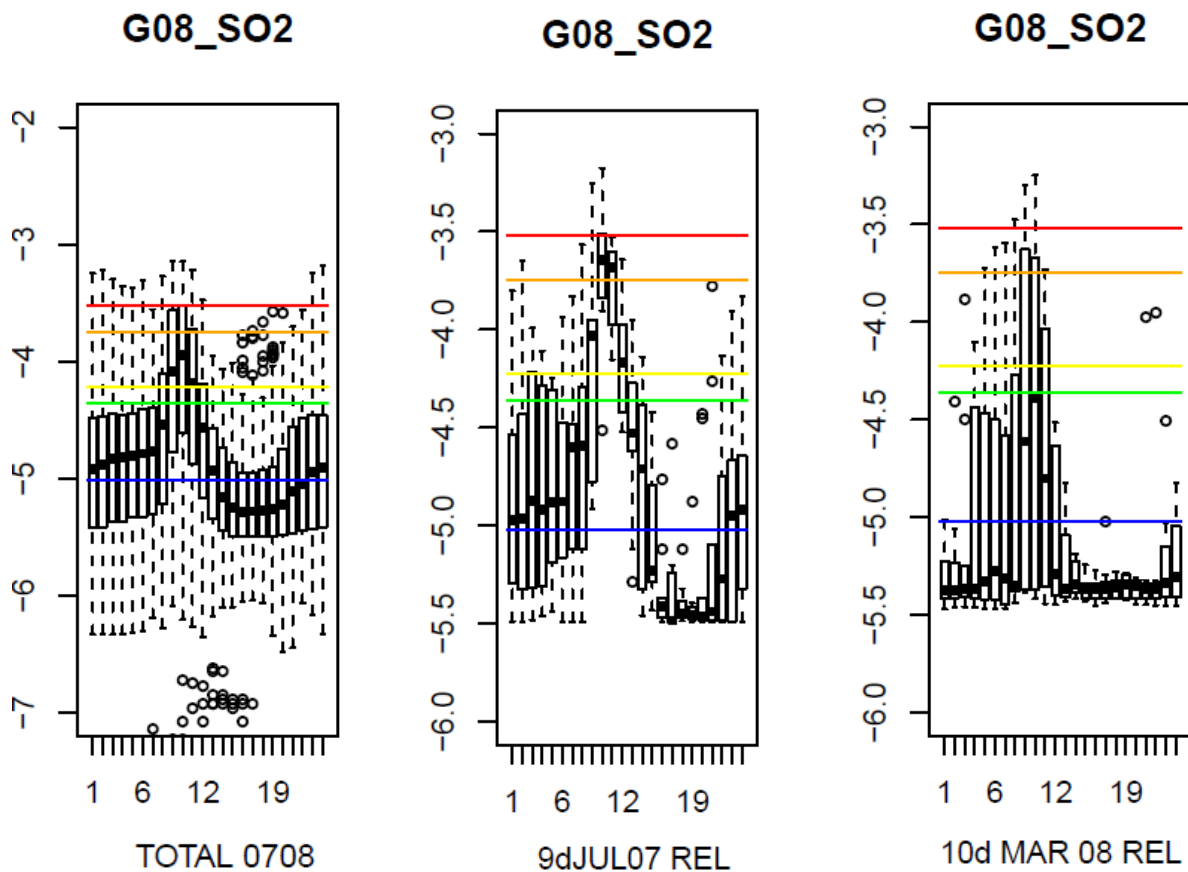


Figure 4.6 - Immission levels ( $1 \times 10^n$ ) vs time (hours) for Station G08 for: Total data (a); July 07 (b); March 08 (c)

Figure 4.6 (a) above shows the hourly data collected by Station G08 during the entire monitoring period of 2007 and 2008, with the darker (black) squares representing the mean value for that hour. It is very similar in shape and trend to stations G01 and G04, either consistently rising then falling. It is at its lowest level at 4pm and steadily rises throughout the evening hours and through the early morning before rising sharply at 8am and triggering the Level 1 alarm at 9am. It maintains this alarm level, reaching its peak SO2 level at 10am, before falling again and leaving the alarm Level at 12pm, where it steadily falls until 4pm.

Figure 4.6 (b) above shows an erratic pattern during the midday but is similar in overall nature to the first graph. The lowest level occurs at 5pm and maintains a

fairly constant horizontal trend line until 10pm where it rises abruptly until 11pm, then rises steadily throughout the night and early morning until 8am where it rises very sharply to trigger the Level 1 alarm at 9am, then the Level 3 alarm at 10am (peak SO<sub>2</sub> level) until after 11am and falling to Level 1 at 12pm, before falling significantly over the next 4 hours to return to its lowest level at 5pm again.

Figure 4.6 (c) above is similar in its general trend to the first graph, but with much lower ambient levels and hourly variation outwith the midday hours. The midday hours, however, have a much larger and abrupt variation. It is still within the 'safe' levels (i.e. less than the Level 1 limit) typically so this station should not be an alarm concern in the summer months.

Station G09 – Huaynacancha

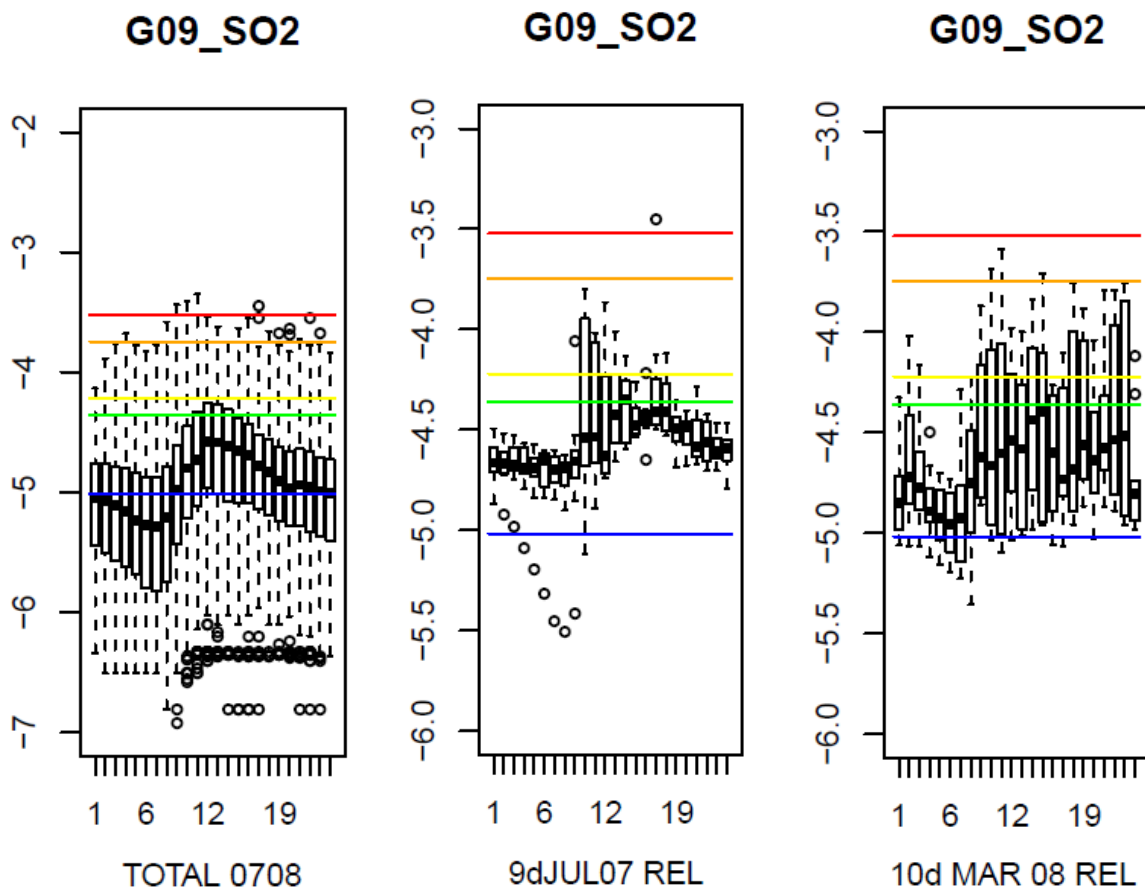


Figure 4.7 - Immission levels ( $1 \times 10^0$ ) vs time (hours) for Station G09 for: Total data (a); July 07 (b); March 08 (c)

Figure 4.7 (a) above shows the hourly data collected by Station G09 during the entire monitoring period of 2007 and 2008. It shows that during a typical day, like the previous examples, the SO<sub>2</sub> levels peak in the midday (in this case, 12pm) and then steadily fall throughout the evening and early morning before rising again (from 8am) and peak at the midday.

Figure 4.7 (b) above shows a very similar pattern to the first (though far less data points and variance), steadily rising through morning to midday before steadily falling through the night again, though it is offset by one or two hours. Following a gentle fluctuation overall, like a wave, with late night & early morning immissions being constant and low level (though the ambient level is higher than the previous graph).

Figure 4.7 (b) above shows a similar pattern to the first for only the early morning (falling steadily) and midday hours (rising steadily). After about 3pm (the peak) the level falls abruptly until 4pm, and falls again until 5pm, where it then begins to steadily rise again until 11pm. It then falls again at midnight and continues through the early morning. Though the levels are all below the limits on average, therefore within 'safe' levels, it is still important to understand the nature of the immissions behaviour for each station.

## **4.2. Results: Bayesian Network Modelling**

The aim was to investigate a possible connection between morning wind patterns and any instances of alarm sounding for each station, as most stations follow the "midday peak" trend as discussed in the qualitative analysis section previously.

Using data collected during multiple days of uninterrupted hourly data, with at least a two-day instance per month (most were longer than two days) across the entire data collection period, the wind pattern categories (REF) or 'signatures' were recorded (or defined) for the hours of 6am through to 11am. These patterns were then compared against each monitoring station and whether that station registered an alarm for that day, ranked in order of alarm level.

The following results serve to show how the model evolved and explain the methodology through real examples in varying levels of complexity.

### **4.2.1. Explanation of data (alarm and wind pattern categories)**

Before interpreting the modelling results, it is important to first understand which data are being used and in what way. The hourly data across the entire monitoring period (July 2007 to Sept 2008, inclusive) has been reduced to two or three (or more if it is available) days in each month with consecutive hourly data.

These data include the SO<sub>2</sub> levels at each hour, which was then reduced to a three hour average for the purpose of identifying which alarm level it came under. If the three hour average is below 500 µg/m<sup>3</sup> there is no alarm (Alarm category D); if the three hour average is between 500 µg/m<sup>3</sup> and 1500 µg/m<sup>3</sup> there is a Level 1 alarm (Alarm category C); if the three hour average is between 1500 µg/m<sup>3</sup> and 2500 µg/m<sup>3</sup> there is a Level 2 alarm (Alarm category B); and if the three hour average is above 2500 µg/m<sup>3</sup> there is a Level 3 alarm (Alarm category A).

This is shown in tabular form below as part of the explanation of the initial model and test.

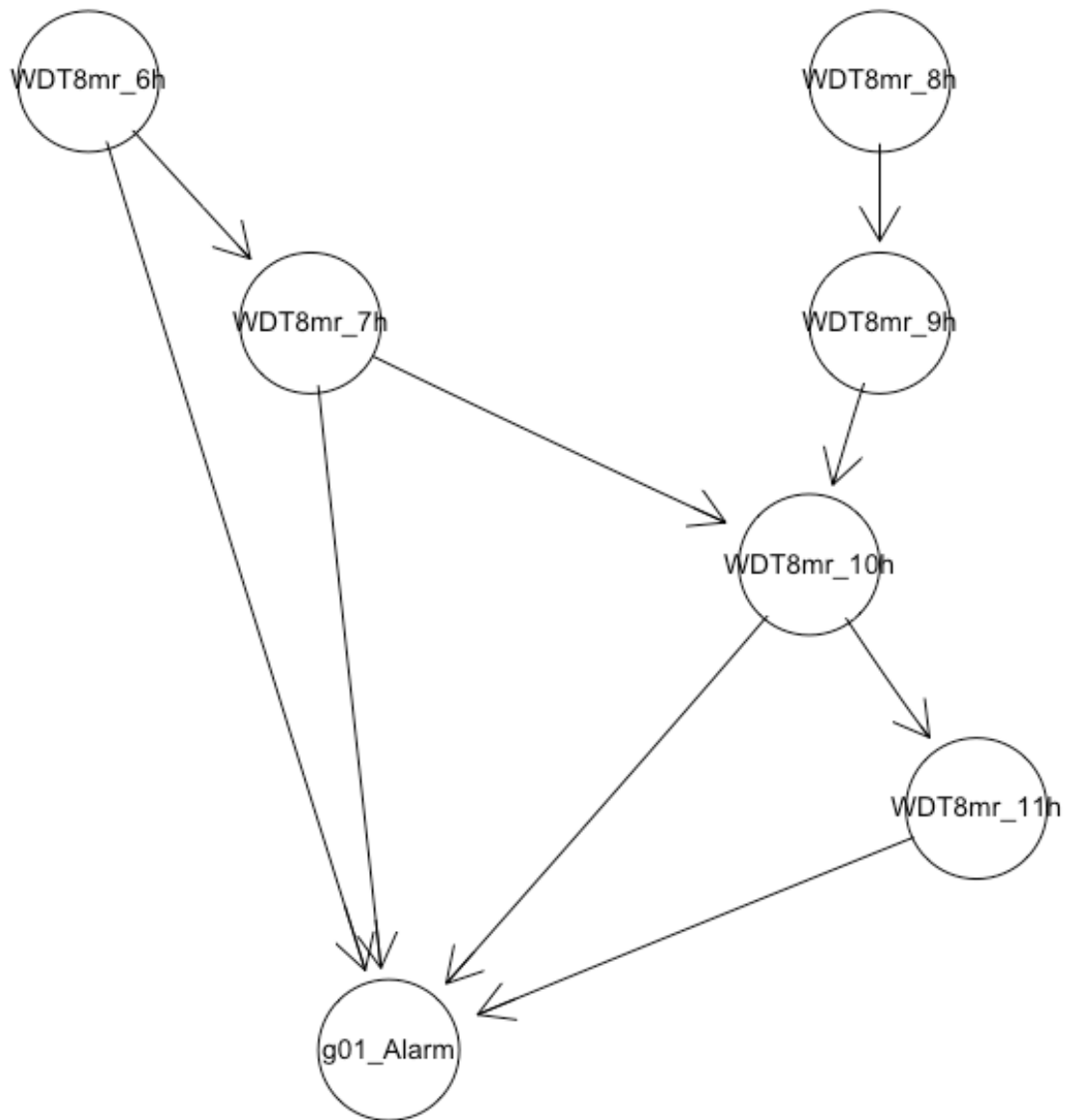
#### 4.2.2. Testing the model: One station

Initially using only some days, one station was checked for alarms and then cross-referenced with wind patterns.

**Table 4.1 - Model test for 10 days in July using one station**

G01									
Year	Month	Day	Highest alarm lvl	Wind Patterns per hour (WDT8mr)					
				6:00 AM	7:00 AM	8:00 AM	9:00 AM	10:00 AM	11:00 AM
2007	July	10	D	71	78	71	72	72	72
2007	July	11	C	71	71	72	71	72	52
2007	July	12	C	52	52	52	52	12	12
2007	July	13	C	71	71	71	72	52	12
2007	July	14	D	71	78	77	77	11	14
2007	July	15	C	71	72	71	71	12	12
2007	July	16	D	77	77	77	71	71	71
2007	July	17	B	57	57	51	71	71	71
2007	July	18	A	77	57	71	71	72	14
2007	July	19	B	51	71	77	71	72	12
2007	July	20	C	71	54	12	12	12	12

Table 4.1 above shows the initial testing period for the Bayesian Model to check for a relationship between morning wind patterns and alarm level. Data was selected from multiple days of consecutive hourly data, in this case eleven days in July 2007. From the table, it can be seen that there is a fairly even spread of alarm levels for G01 over these days, along with some recurring morning winds. This makes for a simple example, with only one station to compare alarm levels and wind patterns, to test the effectiveness of the model and allows the user to check the results intuitively.



**Figure 4.8 - Testing the model using hourly wind and one station's alarms**

Figure 4.8 above shows the graphical results of the initial modelling attempt for G01 alarms vs wind patterns, showing which categories (denoted by the nodes) influence the others (denoted by the connecting lines) and which way the influence lies (the direction of the arrow denotes the influencer). The simplest way to follow the structure is to begin with nodes with no "parent" nodes, ie nodes that influence other nodes without being influenced themselves.

### Summary of results (in chronological order)

- The wind pattern at 6am (WDT8mr\_6h) influences the G01 alarm and influences the wind pattern at 7am (WDT8mr\_7h).
- The wind pattern at 7am (WDT8mr\_7h) influences the G01 alarm and influences the wind pattern at 10am (WDT8mr\_10h).
- The wind pattern at 8am (WDT8mr\_8h) influences the wind pattern at 9am (WDT8mr\_9h).
- The wind pattern at 9am (WDT8mr\_9h) influences the wind pattern at 10am (WDT8mr\_10h).
- The wind pattern at 10am influences alarm G01 and influences the wind pattern at 11am (WDT8mr\_11h).
- The wind pattern at 11am (WDT8mr\_11h) influences alarm G01.

Table 4.2- Summary table of nodes and connections for Figure 4.7

Node	Influences	Influenced by
WDT8mr_6h	WDT8mr_7h; G01 Alarm	None
WDT8mr_7h	WDT8mr_10h; G01 Alarm	WDT8mr_6h
WDT8mr_8h	WDT8mr_9h	None
WDT8mr_9h	WDT8mr_10h	WDT8mr_8h
WDT8mr_10h	WDT8mr_11h; G01 Alarm	WDT8mr_7h, WDT8mr_9h
WDT8mr_11h	G01 Alarm	WDT8mr_10h
G01 Alarm	None	WDT8mr_6h; WDT8mr_7h; WDT8mr_10h; WDT8mr_11h

### Interpretation

This graphic shows that the factors that directly influence the alarm status of station G01 are wind patterns at 6am, 7am, 10am, and 11am. The model shows that the wind at 6am is not dependent on/influenced by anything, however influences both the next wind pattern at 7am and also has an influence on the G01 alarm, so is a “Parent node”.

The probability that the wind at 6am (WDT8mr\_6h) will be a certain pattern is shown in Figure 4.9 overleaf:

## Conditional Probabilities

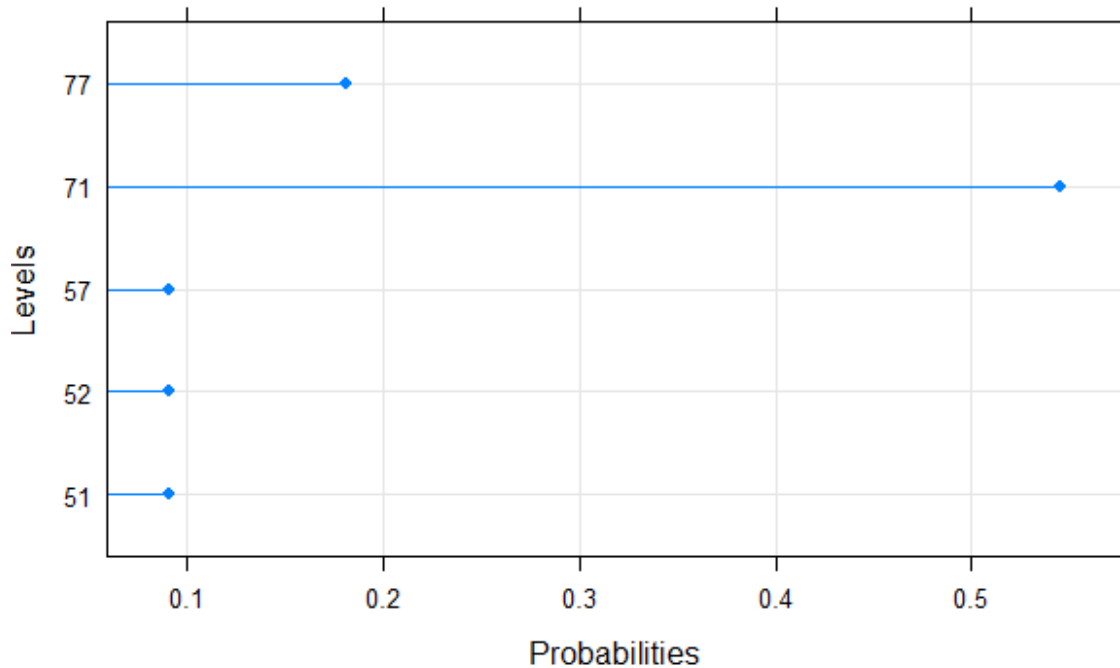


Figure 4.9 - Probabilities of wind patterns for 6am winds in July 11 day testing period

This shows the breakdown of the potential wind patterns at 6am, and their probabilities. This shows that the most common wind pattern at this time is (WP category) 71 with a probability of 55%. The next is WP 77 with 18%, and WPs 57, 52, 51 are equal with roughly 9%.

As this is the first example with the least amount of data, this is not indicative of the entire nature of the winds for the entire year. But it allows the model structure to be developed and checked. As there is only 11 days of data, these probabilities can easily be checked by hand using the initial data, or Table 4.1:

- At 6am, the wind pattern (WDT8mr\_6h) is 77 for 2 days out of 11, or 18%.
- At 6am, the wind pattern (WDT8mr\_6h) is 71 for 6 days out of 11, or 55%.
- At 6am, the wind pattern (WDT8mr\_6h) is 57 for 1 day out of 11, or 9%.
- At 6am, the wind pattern (WDT8mr\_6h) is 52 for 1 day out of 11, or 9%.
- At 6am, the wind pattern (WDT8mr\_6h) is 51 for 1 day out of 11, or 9%.

This can then be done for the other nodes, though as the others (excepting 8am) have influences/dependencies; their probabilities are more complicated as there is a probability table for each previous wind pattern.



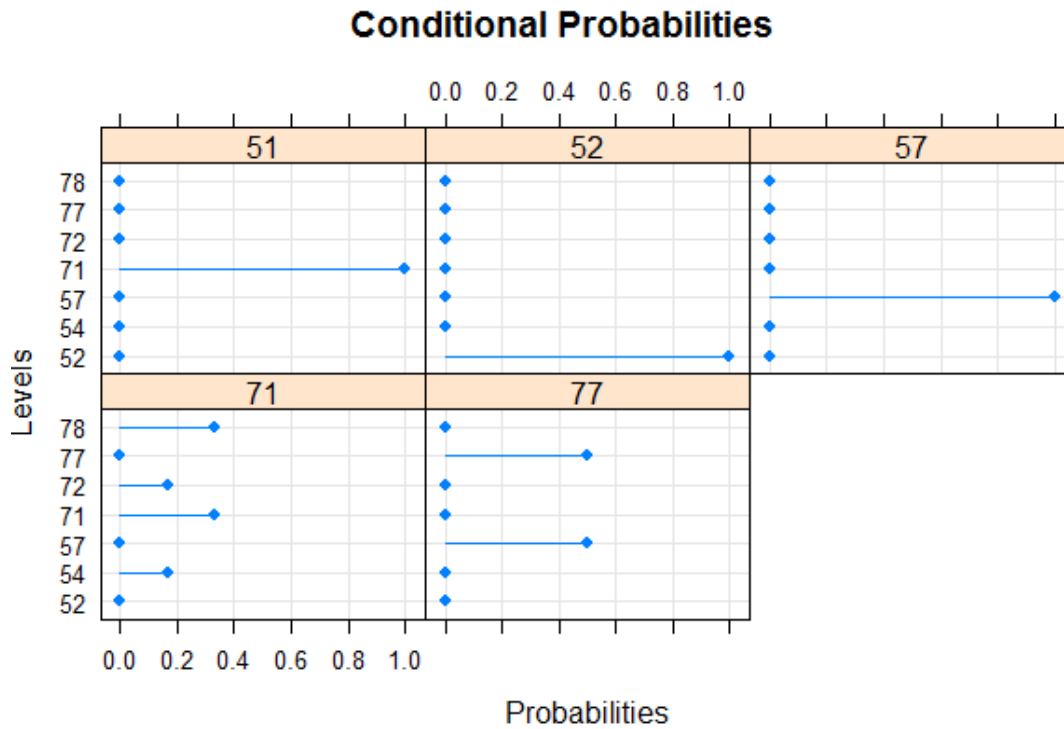


Figure 4.10 - Probabilities of wind patterns for 7am winds in July 11 day testing period based on patterns at 6am (orange bar above)

The figure above shows the probabilities of the wind patterns at 7am (WDT8mr\_7h) based on the wind pattern that occurs at 6am which is the influencer for this node. To explain:

- When the wind pattern at 6am is 51, the wind pattern at 7am has a 100% chance of being 71.
- When the wind pattern at 6am is 52, the wind pattern at 7am has a 100% chance of being 52.
- When the wind pattern at 6am is 57, the wind pattern at 7am has a 100% chance of being 57.
- When the wind pattern at 6am is 71, the wind pattern at 7am could be 78 (34% chance); 72 (16% chance); 71 (34% chance) or 54 (16% chance).
- When the wind pattern at 6am is 71, the wind pattern at 7am has a 50-50 chance of being 57 or 77.

Now that the rationale behind the probability check is understood, the following figures show the probabilities for the other nodes:

### Conditional Probabilities

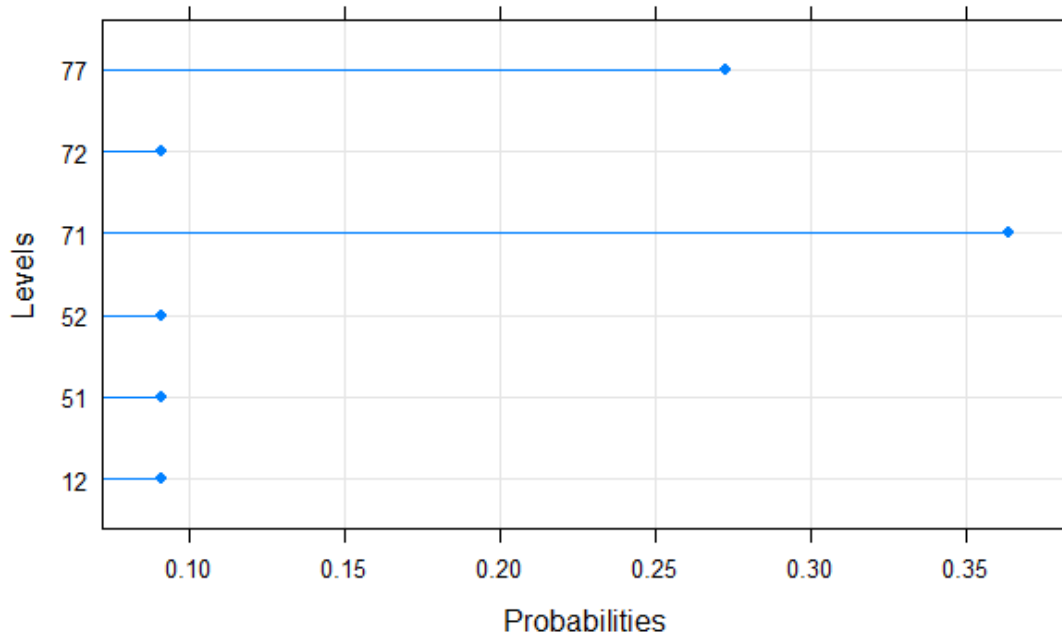


Figure 4.11 - Probabilities of wind patterns for 8am winds in July 11 day testing period

### Conditional Probabilities

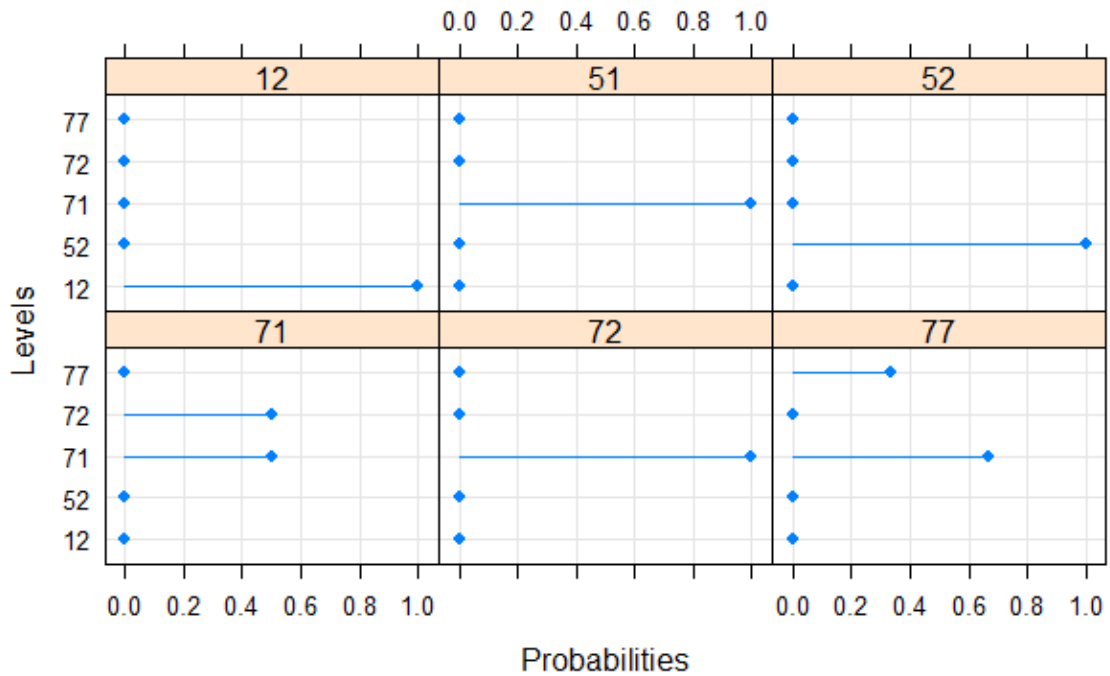


Figure 4.12 - Probabilities of wind patterns for 9am winds in July 11 day testing period based on patterns at 8am (orange bar above)

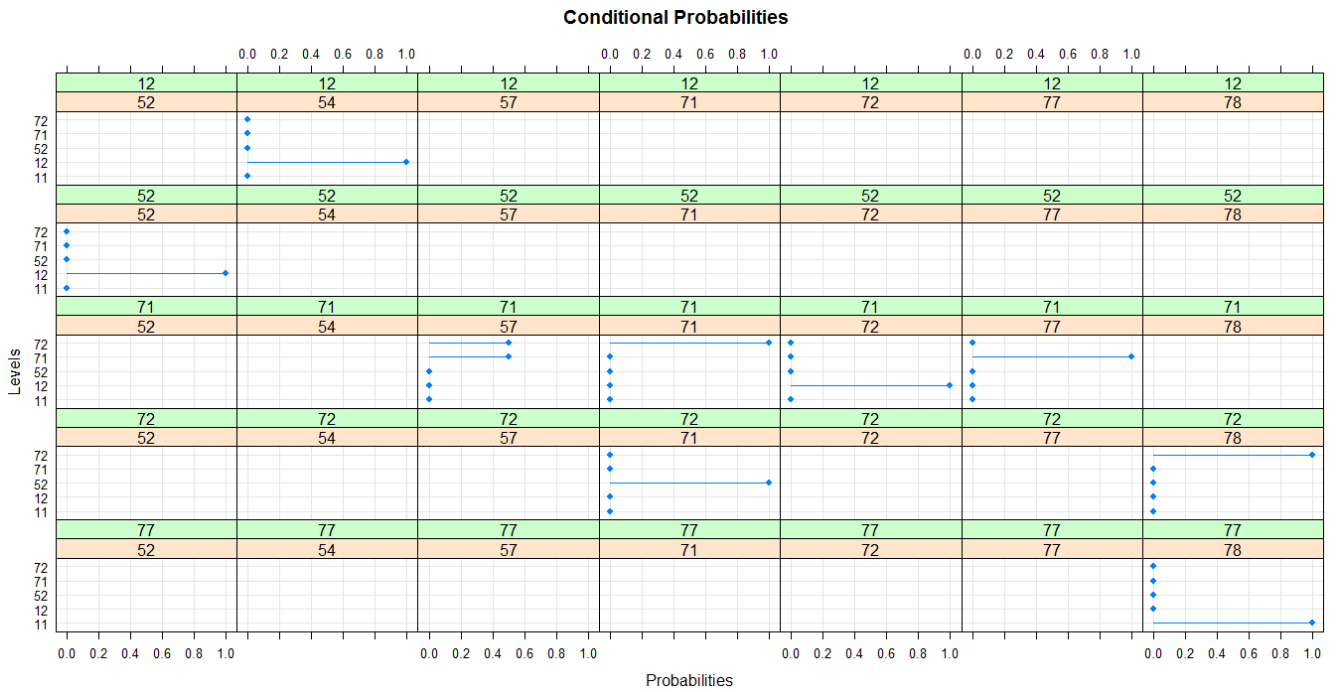


Figure 4.13 - Probabilities of wind patterns for 10am winds in July 11 day testing period based on patterns at 7am (orange bar above) & 9am (green bar above)

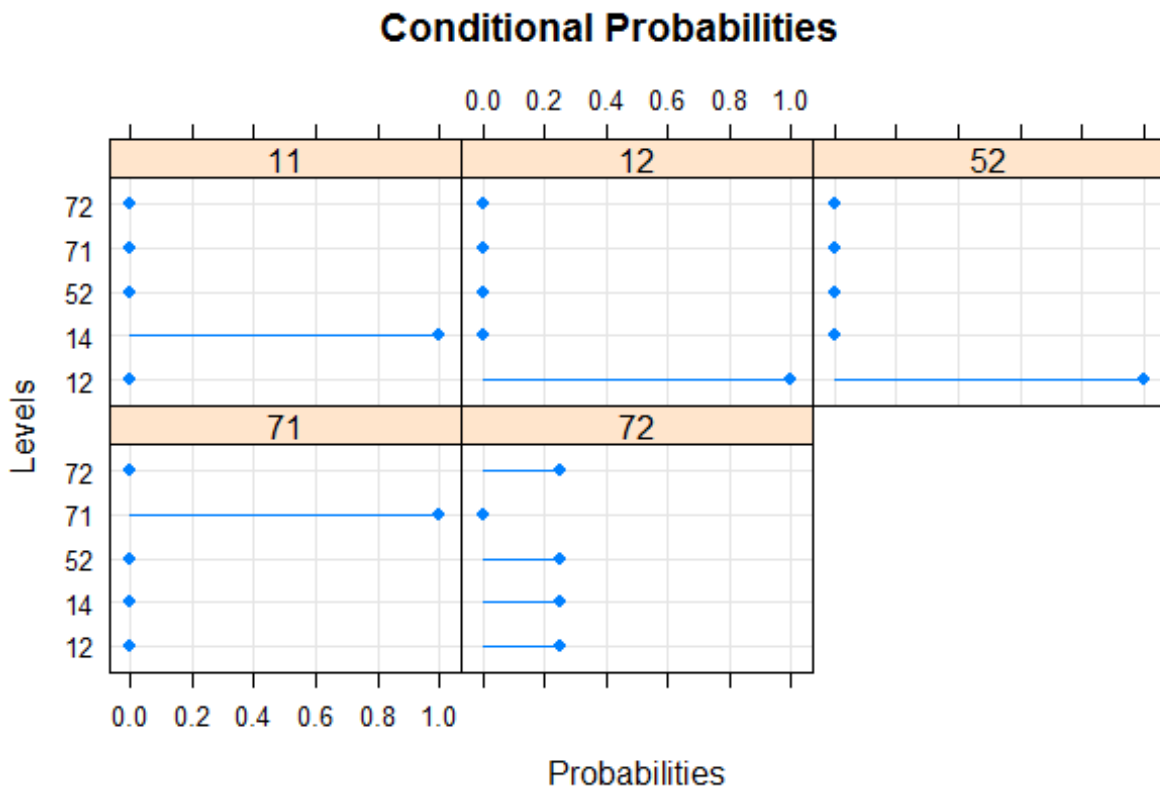


Figure 4.14 - Probabilities of wind patterns for 11am winds in July 11 day testing period based on patterns at 10am (orange bar above)

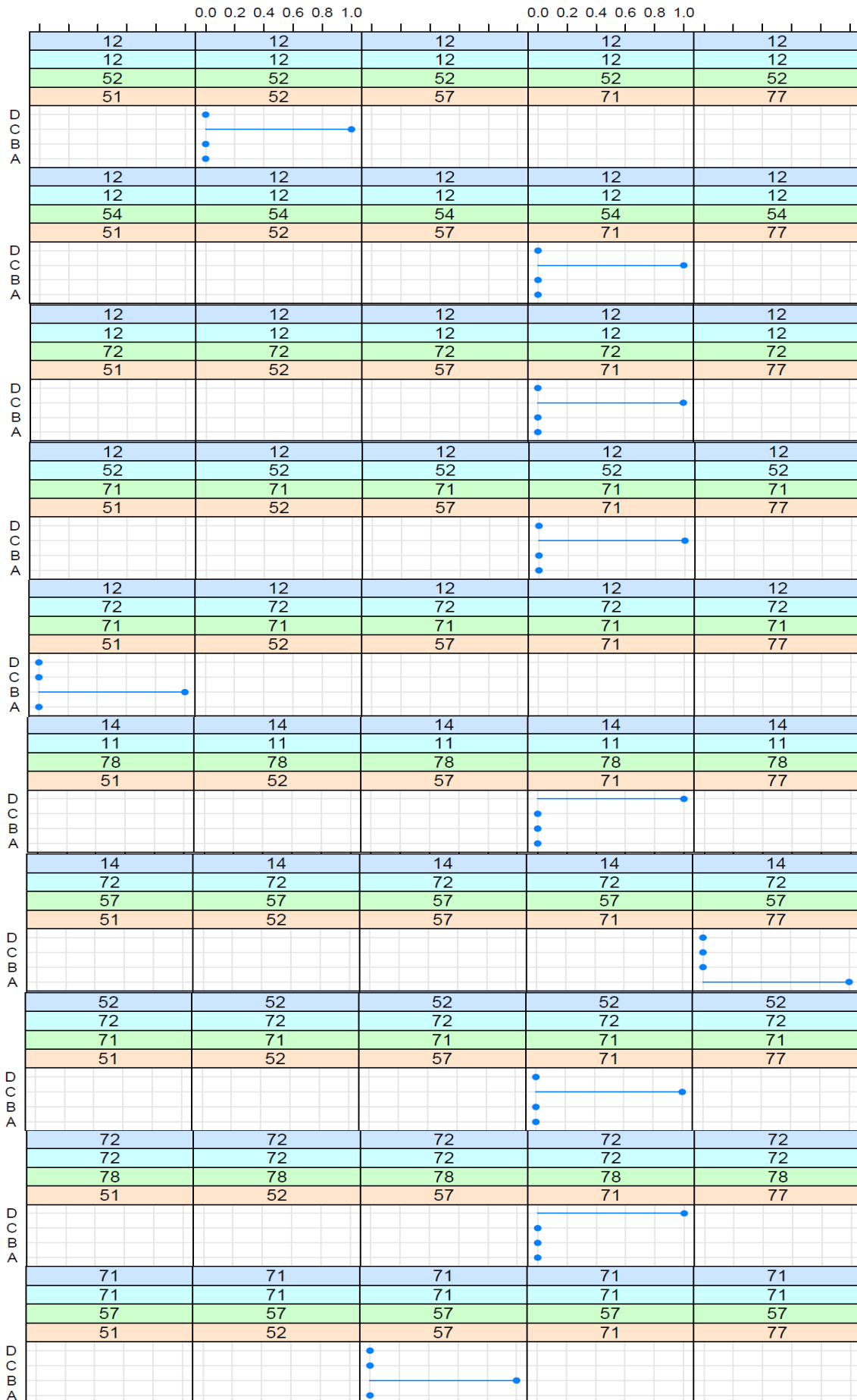


Figure 4.15 - Probabilities alarm level for G01 in July 11 day testing period based on wind patterns at 6am (orange bar above); 7am (green bar above); 10am (cyan bar above); 11am (blue bar above)

Lastly, after the model has returned probabilities for the relationships between wind patterns, it can be checked for the probabilities of an alarm sounding. As there are so many pattern combinations to check through (as every WP must be combined with every other WP), where many 'potential' combinations do not appear in practice, only the combinations that returned results have been included. This is shown in Figure 4.12 on the previous page.

It shows the alarm levels triggered during the 11 day testing period and the morning wind patterns that had been involved. The number denoted by the orange bar refers to the wind pattern classification at 6am; the green bar refers to the wind pattern at 7am; the cyan (light blue) bar refers to the wind pattern at 10am; and the blue bar (darker blue) refers to the wind pattern at 11am.

These results show that modelling the alarm levels against the wind pattern based on empirical data can be done, and that many more days must be included (alongside the other monitoring stations and their alarms) to reach some level of effectiveness.

The following sections show how the model grew with the addition of the other stations, more days (first for July only, and then for a few days of each month), before reaching a point where it can be considered useful. The probabilities are not included for each of these stages, however, as, though they are necessary to explain the model, they are only really relevant at the first stage (to explain the model when it is still simple with relatively few data points) and at the final stage, when it can be discussed what to do with the predictions in the model to improve the situation at La Oroya.

### 4.2.3. Testing the model: More stations

After this, the model was checked with more stations and using more days' worth of data for July. The stations chosen were G01, G03, G05 and G08. These were chosen as they has similar immissions behaviour throughout the day (as shown in "Qualitative Results" previously) and typically had alarm behaviour during the midday hours only, thus were deemed to be similar in nature.

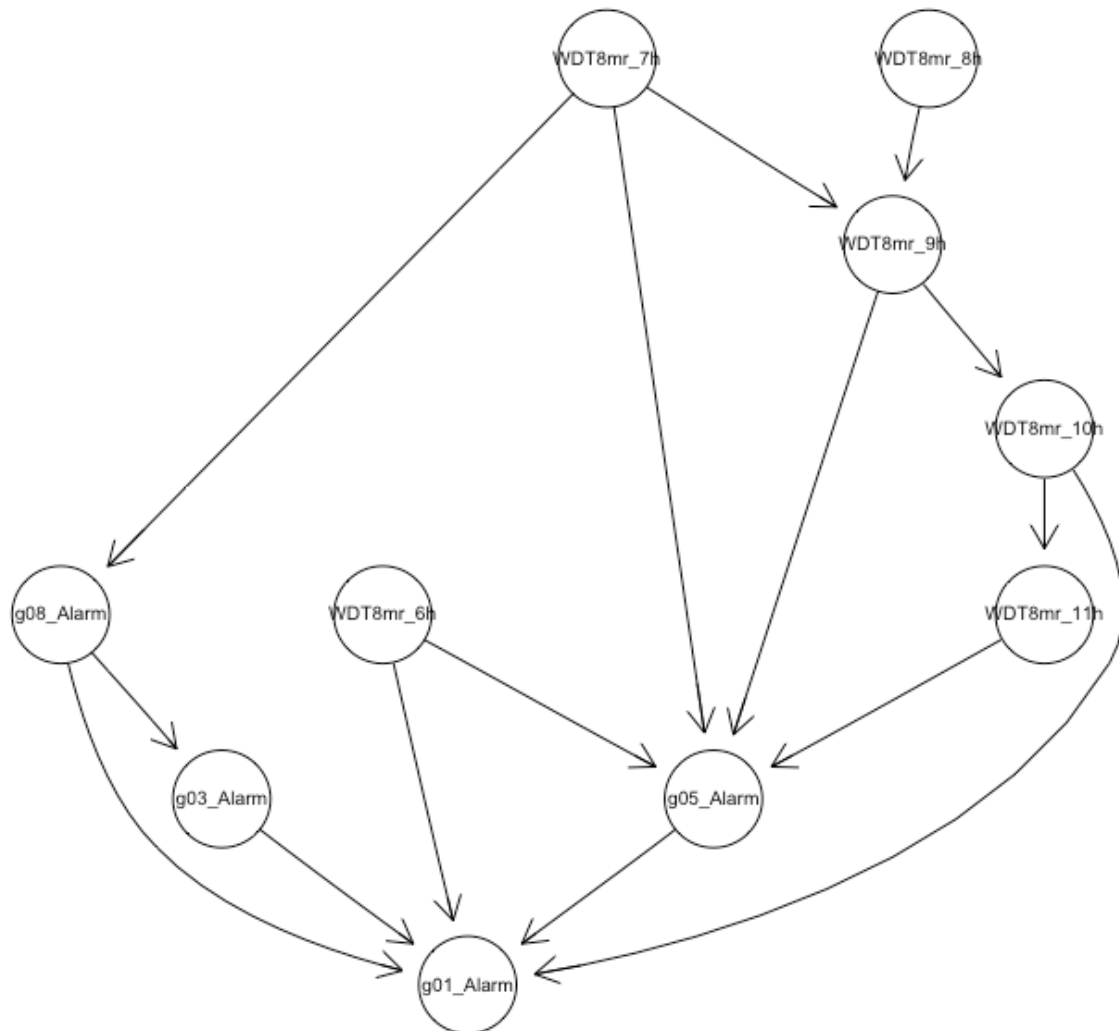


Figure 4.16 – Testing the model with more stations (G01, G03, G05, and G08)

Figure 4.16 above shows the graphical results of the model for G01, G03, G05 and G08 alarms vs wind patterns, showing some similar connections as before, as well as some changes and some completely new connections due to the addition of other stations and how their inclusion affects the model.

## Summary of nodes:

The addition of more available data has shown that, for **wind patterns**:

- The wind pattern at 6am (WDT8mr\_6h) no longer has an impact on the wind at 7am (WDT8mr\_7h), though it retain its connection to the G01 alarm directly, and is also found to be connected to the G05 alarm level. This node remains a “Parent node”.
- The wind pattern at 7am (WDT8mr\_7h) no longer has a direct relationship with the alarm at 10am (WDT8mr\_10h) as before, but is now found to be directly related to the WP at 9am (WDT8mr\_9h). It is also no longer related to the alarm at G01, but instead the alarms at G05 and G08.
- The wind pattern at 8am (WDT8mr\_8h) retains its connection with the WP at 9am (WDT8mr\_9h). It also remains a “parent node” and is still uninfluenced by wind patterns before.
- The wind pattern at 9am (WDT8mr\_9h) retains its connection with the WP at 8am (WDT8mr\_8h) and is now also affected by the WP at 7h (WDT8mr\_7h). It is found to have an influence on the alarm level at station G05.
- The wind pattern at 10am (WDT8mr\_10h) retains its connection with the WP at 11am (WDT8mr\_11h). And is now only influenced by the WP at 9am (WDT8mr\_9h), and no longer influenced by the WP at 7am (WDT8mr\_7h). It retains its connection with the alarm at G01.
- The wind pattern at 11am (WDT8mr\_11h) retains its connection with the WP at 10am (WDT8mr\_10h). It no longer influences the alarm at G01, and is instead found to influence the alarm at station G05.

The addition of more available data has shown that, for **station alarms**:

- The alarm level at station G01 is directly influenced by the wind patterns at 6am (WDT8mr\_6h) and 10am (WDT8mr\_10h), and by the alarm level at stations G03, G05, and G08.
- The alarm level at station G03 is directly influenced by the alarm level at G08. It has a direct influence on the alarm level at station G01.
- The alarm level at station G05 is directly influenced by the wind patterns at 6am (WDT8mr\_6h), 7am (WDT8mr\_7h), 9am (WDT8mr\_9h), and 11am (WDT8mr\_11h). It has a direct influence on the alarm level at station G01.
- The alarm level at station G08 is directly influenced by the wind pattern at 7am (WDT8mr\_7h). It has a direct influence on the alarm level at stations G03 and G01.

#### 4.2.4. Testing the model: All days

The model was then checked using data from all available days – with consecutive hour data collection and no gaps – throughout the year. This served to provide an insight into the month to month changes and how they fit into the larger, seasonal narrative. Thus, a “Month” node was added to determine if a particular month can also have an impact on the final results of an alarm sounding.

The number of stations used was also increased, involving every station bar G04 as, due to its alarm-heavy nature, it was found to be “eclipsing” the connections between the other stations, i.e. it is at the highest alarm level so often and for so long that it skews the results of the other stations as their connections are deemed ‘weaker’ in comparison.

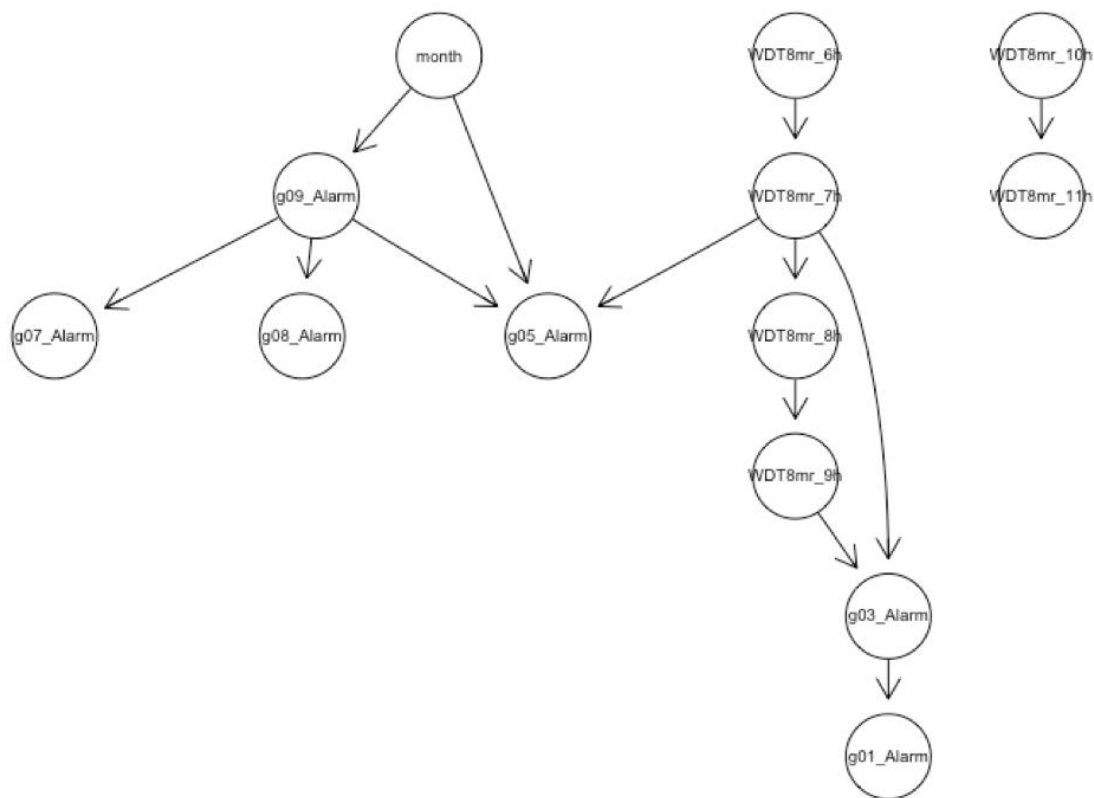


Figure 4.17 – Testing the model with all days and stations (except G04)

Figure 4.17 above shows that the wind patterns at each hour directly influence each other, and serve to form the ‘spine’ of the model and it would be expected that the alarm nodes would be based on this, and that their nodes would surround the wind pattern core of the model.

However, this is not the case with Figure 4.17. The alarm nodes on the left are either solely based on a month (G09) or based on the level of another alarm (G07 and G08 based on G09). While an alarm level at one station can be an indicator for another, as from the data there are certainly stations that tend to sound similar alarms at



similar times (e.g. G01 and G03 tend to both sound Alarm Level 1 within an hour or two of each other), it is not the whole story. To assume an alarm level at one station solely based on an alarm level at another, and not take the wind pattern into consideration, would be an oversight. After all, it is the wind that carries the pollutant that causes the alarm. An alarm level prediction should be based on a combination of both the wind pattern and the alarm level at its connected station.

This result, though using all the data, was considered to be too simple with regards to the connection between wind patterns and alarms. More testing is needed using data with a broader range of alarm presence and variability, even if it must be with fewer days.

#### 4.2.5. Testing: More alarm options

The model structure was checked using data from February through April, as there is more alarm variability in these months. The reduction in available days, and therefore wind pattern information, mean that the model will not present the wind patterns as accurately. However, it will showcase the effects the hourly wind patterns have on alarm level more than the model using all the data. It may be more inaccurate overall, but if it shows a truer representation of the wind to alarm structure, it will be useful.

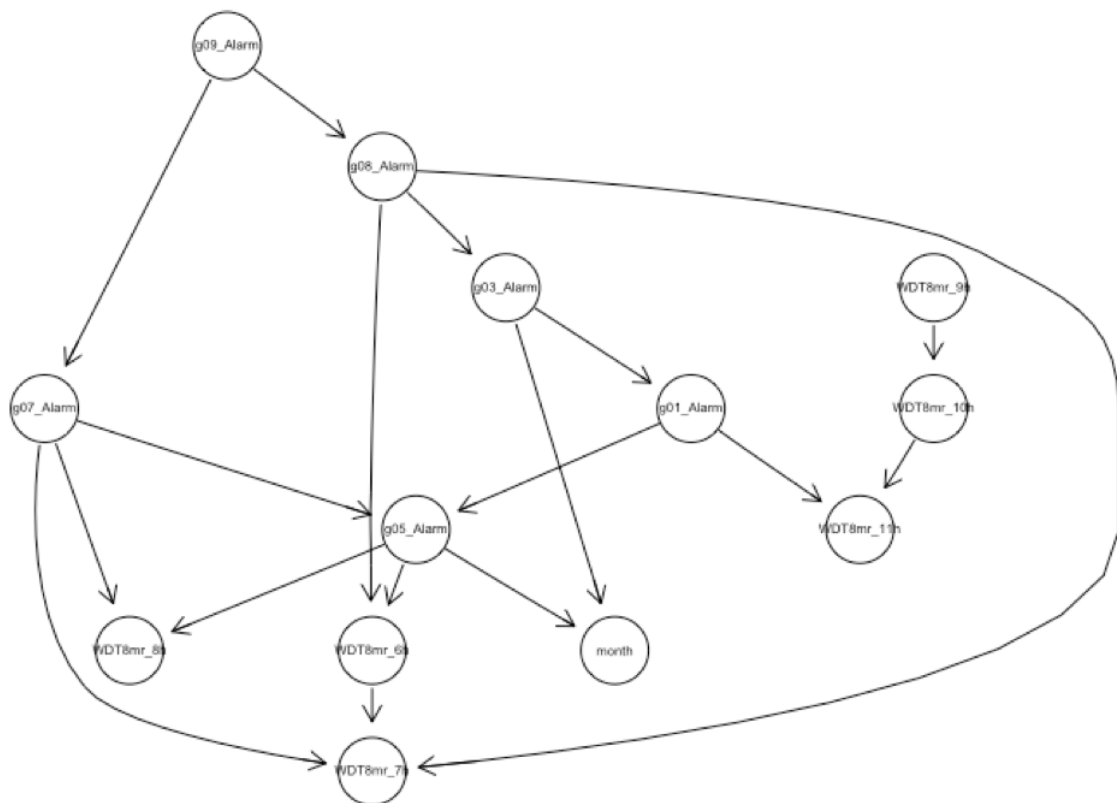


Figure 4.18 – Testing the model for Feb-Apr with all stations (except G04)

As predicted, using reduced days has caused the wind pattern connections to regress to a partially disconnected scatter of nodes, similar to the first few model tests. This is due to having fewer available days and makes these patterns less apparent. The model does have more alarm variability however, and this is reflected in the increased number of connections between WPs and alarms.

Something must also be said for this model regarding the orientation of the connecting lines. For example, the alarms for G03 and G05 are both found to be influencing the “month” node, though logically this would be the reverse. In this way, expert knowledge must be applied to the connections highlighted by the model to ensure it is logically sound.

#### 4.2.6. Final Model

Using the data of all the days to define the structure with regards to wind behaviours and combining it with the alarm structure from the alarm variability in months 2 to 4, a mixed network has been developed for the final model. This is shown in Figure 4.19 below:

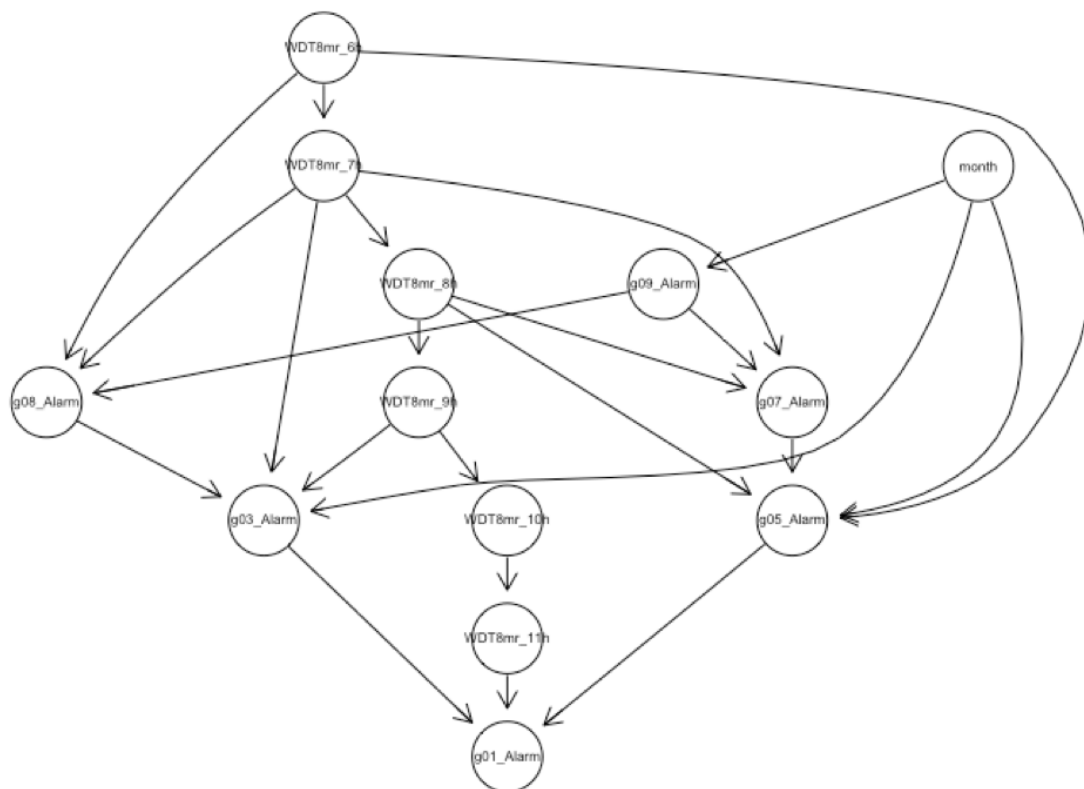


Figure 4.19 – Final model structure

This represents the intuitive assumption that wind behaviour at one hour has an effect on the wind behaviour of the next hour. The addition of the structure for the stations’ alarm levels due to: using the months with greater availability of alarms; and refining the “month” node which represents the influence of monthly variability and alludes to seasonal differences in wind behaviour and air flow, therefore in alarm presence.

## Summary of nodes:

The addition of more available data has shown that, for **wind patterns**:

- The wind pattern at 6am (WDT8mr\_6h) is a “parent node” and is not influenced by any other node in the model. It directly influences the wind pattern at 7am (WDT8mr\_7h) and the alarms at stations G05 and G08.
- The wind pattern at 7am (WDT8mr\_7h) is directly influenced by the wind pattern at 6am (WDT8mr\_6h). It directly influences the wind pattern at 7am (WDT8mr\_7h) and the alarms at stations G05, G07, and G08.
- The wind pattern at 8am (WDT8mr\_8h) is directly influenced by the wind pattern at 7am (WDT8mr\_7h). It directly influences the wind pattern at 9am (WDT8mr\_9h) and the alarms at stations G05 and G07.
- The wind pattern at 9am (WDT8mr\_9h) is directly influenced by the wind pattern at 8am (WDT8mr\_8h). It directly influences the wind pattern at 10am (WDT8mr\_10h) and the alarm at station G03.
- The wind pattern at 10am (WDT8mr\_10h) is directly influenced by the wind pattern at 9am (WDT8mr\_9h). It directly influences the wind pattern at 11am (WDT8mr\_11h) only.
- The wind pattern at 11am (WDT8mr\_11h) is directly influenced by the wind pattern at 10am (WDT8mr\_10h). It directly influences the alarm at station G01.

The addition of more available data has shown that, for **alarm levels**:

- The alarm at G01 is directly influenced by the wind pattern at 11am (WDT8mr\_11h) and the by the level of alarms at stations G03 and G05.
- The alarm at G03 is directly influenced by the wind patterns at 7am (WDT8mr\_7h) and 9am (WDT8mr\_9h), by the level of alarms at station G08, and by the “month” node. It directly influences the alarm level at station G01.
- The alarm at G05 is directly influenced by the wind patterns at 6am (WDT8mr\_6h) and 8am (WDT8mr\_8h), by the level of alarms at station G07, and by the “month” node. It directly influences the alarm level at station G01.
- The alarm at G07 is directly influenced by the wind patterns at 7am (WDT8mr\_7h) and 8am (WDT8mr\_8h) and by the level of alarm at station G09. It directly influences the alarm level at stations G05.
- The alarm at G08 is directly influenced by the wind patterns at 6am (WDT8mr\_6h) and 7am (WDT8mr\_7h) and by the level of alarm at station G09. It directly influences the alarm level at station G07.
- The alarm at G09 is directly influenced by the month. It directly influences the alarm level at stations G07 and G08.

Lastly, the month node is a “parent node” which is logical as the months are not dependant on anything. The month node has a direct influence on the alarm levels at stations G03, G05, and G09.

**Example of model in use, with conditional probabilities:**

- Model output for G01 & G03 alarms if the wind pattern at 6am is 71:

**Table 4.3 - Alarm probabilities for G01 & G03 based on 6am WP71**

		G01_Alarm			
		A	B	C	D
G03_Alarm	A	0.02403846	0.03846154	0.01923077	0.00000000
	B	0.03846154	0.04807692	0.00000000	0.00000000
	C	0.00000000	0.00000000	0.66346154	0.08653846
	D	0.00000000	0.00000000	0.00000000	0.08173077

If the wind pattern at 6am is 71, the model returns a 66.3% probability that the alarms at G01 and G03 will be category C. The next highest probability is an 8.6% chance that station G01 will be category D (no alarm) and G03 will be category C. There is an 8.1% chance that G03 will be category D and have no alarm.

- Model output for G01 & G03 alarms if the WP at 6am is 77:

**Table 4.4 - Alarm probabilities for G01 & G03 based on 6am WP77**

		G01_Alarm			
		A	B	C	D
G03_Alarm	A	0.10318949	0.12945591	0.05628518	0.00000000
	B	0.03564728	0.03564728	0.03752345	0.00187617
	C	0.00000000	0.00000000	0.27579737	0.17823640
	D	0.00000000	0.00000000	0.01500938	0.13133208

If the wind pattern at 6am is 77, the model returns a 27.6% probability that the alarms at G01 and G03 will be category C. The next highest probability is a 17.8% chance that station G01 will be category D (no alarm) and G03 will be category C.

- Model output for G01 & G03 alarms if the WP at 6am is 77; at 7am is 57; at 8am is 57; and the month is July:

Table 4.5 - Alarm probabilities for G01 & G03 based on 6am WP77; 7am WP57; 8am WP57 in July

		G01_Alarm			
		A	B	C	D
G03_Alarm	A	0.03389831	0.02542373	0.04237288	0.00000000
	B	0.00423729	0.02542373	0.05508475	0.01271186
	C	0.00000000	0.00000000	0.46610170	0.33474576
	D	0.00000000	0.00000000	0.00000000	0.00000000

This returns a 46.6% probability that the alarms at G01 and G03 will be category C. The next highest probability is a 33.5% chance that station G01 will be category D (no alarm) and G03 will be category C. There is a 3% chance they will both be category A and a 2.5% chance they will both be B. There is also a 0% chance that G03 will have no alarm (category D).

- Model output for G01 & G03 alarms if the WP at 6am is 57; at 7am is 57; at 8am is 77; at 9am is 71; at 11am is 72; and in the month of July:

Table 4.6 - Alarm probabilities for G01 & G03 based on 6am WP57; 7am WP57; 8am WP77; 9am WP71; 11am WP 72; month of July

		G01_Alarm			
		A	B	C	D
G03_Alarm	A	0.15254237	0.04495210	0.04863670	0.00000000
	B	0.04053058	0.05305822	0.12748710	0.04347826
	C	0.00000000	0.00000000	0.25423729	0.23507738
	D	0.00000000	0.00000000	0.00000000	0.00000000

This returns a 25.4% probability that the alarms at G01 and G03 will be category C. The next highest probability is a 23.5% chance that station G01 will be category D (no alarm) and G03 will be category C. There is a 15.2% chance they will both be category A. There is a 0% chance that G03 will have no alarm (category D).

### 4.3. Limitations of Data

By the time of the final model, all of the data collected across the two year collection period was used (narrowed down to days of continuous hourly collection). However, some months had more days of uninterrupted data collection than others. For example, the month of July had consistently more data available and suitable for use than other months. This is expressed in the Figure 4.15 below.

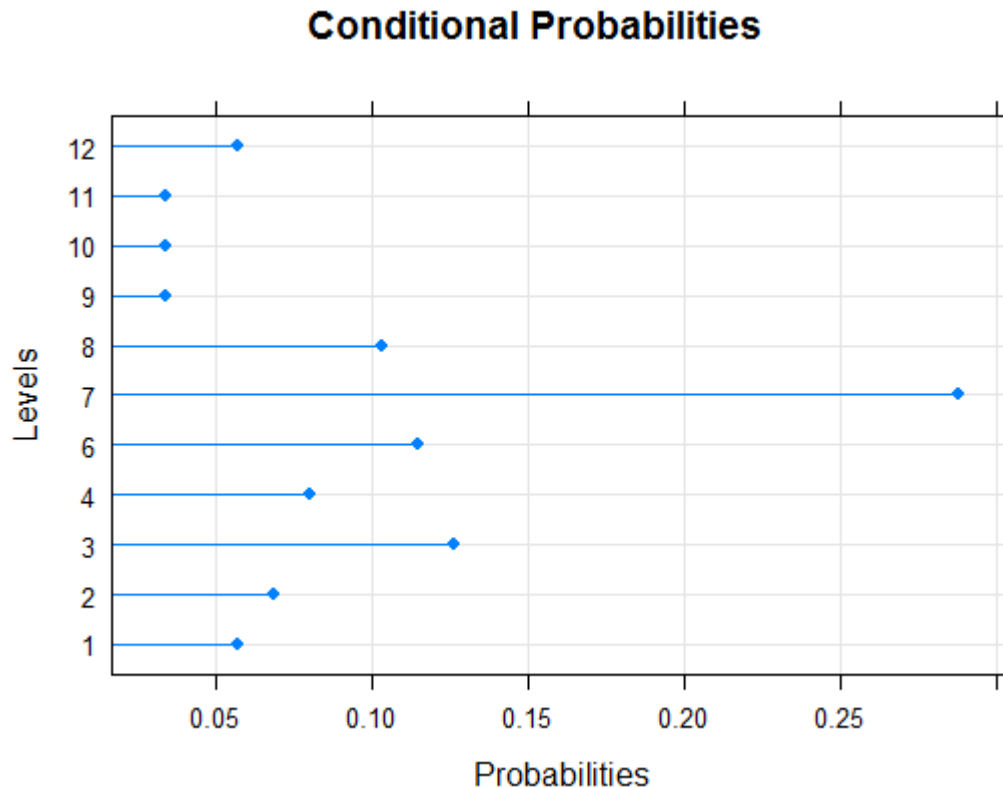


Figure 4.20 – Probability of a data point being sourced from a particular month

So, from Figure 4.20 it can be seen that when comparing all of the available data, around 28% of that data was from the month of June (includes both 2007 and 2008). While this is included in the model as all available data is required to make the model as effective as possible, it is still a disproportionate amount days from just one month. This means that there is an inherent bias/skew in the model and therefore the model will reflect the winter months more accurately than the summer months.

Further work is required with regards to data collection. Or the choice must be made to use non-complete days to take advantage of the abundance of data there (there will always be more non-complete days than days of continuous collection). Until then, the evidence requirements could be relaxed, such as using a probability estimate made up of more than one wind pattern, for example, at 6am consider both WP 71 and WP 72, instead of only one or the other. This could serve to give a decent result in the meantime, while more data is collected and the model becomes more sophisticated.

## 5. Potential Application of the Model

If the model becomes accurate and advanced enough there is potential for application to the day to day operations of the La Oroya metallurgical complex. If more data is added with days of winds and days of station alarm presence/absence then the model will be refined and improve its accuracy at each addition of data and over time these iterations will resemble the true situation more and more i.e. the probabilities for a certain wind pattern occurring or alarm sounding will be true to form.

If the model is built upon, with each iteration it could be made more advanced in its use/interface and therefore be of more use to the operations team at the complex. For example, assuming a high level of accuracy, the person in charge of the daily operations of the complex should be able to use the model at 7am and, based on the wind pattern recorded at 6am, be able to know what the chances were of (a) a certain wind pattern at 8am, and (b) what the chances were for an alarm being raised at a certain station. Then at 8am, based on wind patterns from 6am and 7am, the probabilities will be refined and have more meaning. The same for 9am etc. And once a probability for an alarm is above a certain threshold, action can be taken either in the form of an alert to the people living near that station (within a certain effective radius) with a warning of a potential alarm, or if the risk is high enough (or low risk but high potential effect, such as a level 3 alarm for many stations) then the factory can suspend operations for a few hours and resume later when the wind has changed to a more favourable condition.

This scenario would of course require data collection & analysis automatically and in real time, which would require upgrade/replacement of the monitoring stations to meet the necessary broadcasting infrastructure requirements. The advantage of this however is the fact that data would be collected daily adding to the wealth of empirical data available to use for the model, so the model would be refined and improved daily.

## 6. Conclusion & Recommendations

This study presents a methodology for developing and testing a model to simulate SO<sub>2</sub> pollutant immision alarms, from information collected via monitoring stations in key areas around the town of La Oroya, based on the behaviour of local winds, categorised using previous work in this subject area and location.

This study has shown that the modelling of wind patterns and relating these to local immision alarm level can be done, however more data is required to update the model before it can have any operational use. Therefore the recommendations for further study in this topic firstly involve the collection of more data at the stations, as the more data collected means there is more chance of getting consecutive days of hourly data and will improve the accuracy of the model, and lessen the effects of monthly skew.

Additionally, it is recommended to broaden the focus of this study into other pollutant contaminants, with the logical first choice being PM<sub>10</sub>. The model can either be applied to PM<sub>10</sub> levels to test its effectiveness, or be updated with a PM<sub>10</sub> 'option' as the PM<sub>10</sub> model may be slightly different than the SO<sub>2</sub> prototype.

The framework of this model, as well as the underlying methodology of wind pattern classifications, could also be applied to other areas of complex wind behaviour, like other mountain valleys or canyons etc, or heightened levels of air pollution from industries or transport etc.

The findings presented in this study highlight the situation in La Oroya, Peru, where the main industrial complex provides both a large source of employment but also a large source of atmospheric emissions. Further developing the model presented in this work would allow operational changes to be made to the day to day emissions output of the complex, based on daily wind pattern analyses and immision alarm forecasts, and could be used as part of a wider strategy to ensure the residents of the town no longer having to make the choice between having economic prosperity or a good quality of life.



## 7. References

- Arnfield, A. (2003) *Two decades of urban climate research: A review of turbulence, exchanges of energy and water, and the urban heat island*. Int. J. Climatol., **23**: 1–26, DOI:10.1002/joc.859
- Arroyo, A., Corchado, E., & Tricio, V. (2009) Computational Methods for Immision Analysis of Urban Atmospheric Pollution, Department of Civil Engineering & Department of Physics, University of Burgos, Burgos, Spain. Proceedings of the International Conference on Computational and Mathematical Methods in Science and Engineering, CMMSE2009
- Australian Government: Department of the Environment and Heritage (2005) *Air quality fact sheet*, [Online Resource] Available at: <http://www.environment.gov.au/protection/publications/factsheet-sulfur-dioxide-so2> [Accessed: 16/04/2016]
- British Columbia Ministry of Environment (2015) *A Primer on the British Columbia Air Quality Dispersion Modelling Guideline*. Clean Air Section, Environmental Standards Branch, Environmental Protection Division, British Columbia Ministry of Environment.
- Bellinger, D. & Bellinger, A. (2006) *Childhood Lead Poisoning: The Tortuous Path from Science to Policy*. 116:4 J. Clinical Investigation 853-857
- Chen, X.-L., Zhao, H.-M., Li, P.-X., and Z.-Y. Yin (2006) *Remote sensing image-based analysis of the relationship between urban heat island and land use/cover changes*. Remote Sens. Environ., **104**: 133–146, DOI:10.1016/j.rse.2005.11.016
- Cornejo, A & Gottesfeld, P. (2004) *Interior Dust Lead Levels in La Oroya, Peru*. Lima: Asociacion Civil Labor, Occupational Knowledge International and Cooperation, Lima.
- Darby, L. S (2005) *Cluster analysis of surface winds in Houston, Texas, and the impact of wind patterns on ozone*. Journal of Applied Meteorology **44**: 1788-1806. DOI: 10.1175/JAM2320.1
- D.E. & B.O.C. (2007) Examen Especial al Complejo Metalúrgico La Oroya. Desarrollo y Ecología S.A.C. and B.O. Consulting S.A. for Programa de Adecuación y Manejo Ambiental del Complejo Metalúrgico de La Oroya, Ministerio de Energía y Minas, Republica del Perú. No 002-2007-MEM-DGM/V, Jan 3.
- Drobinski, P., Dabas, A. M., Haeberli, C., and Flamant, P. H., (2003) *Statistical characterization of the flow structure in the Rhine valley*. Boundary-Layer Meteorology 106: 483-505, DOI: 10.1023/A: 1021262321679
- Edgerton, S. (1999) *Particulate air pollution in Mexico City: A collaborative research project*. J. Air Waste Manage. Association, **49**: 1221–1229, DOI:10.1080/10473289.1999.10463915
- Environmental Protection Agency (2010) *Particulate matter and health fact sheet*, [Online Resource] Available at: <https://www3.epa.gov/pm/health.html> [Accessed: 16/04/2016]
- Grimmond, S. (2007) *Urbanization and global environmental change: Local effects of urban warming*. Geogr. J., 173, 83–88, DOI:10.1111/j.1475-4959.2007.232\_3.x
- Haeger-Eugensson, M. & Holmer, B. (1999) *Advection caused by the urban heat island circulation as a regulating factor on the nocturnal urban heat island*. International Journal: Climatol, 19, 975–988, DOI:10.1002/(SICI)1097-0088(199907)19:9<975::AID-JOC399>3.0.CO;2-J

- Janhall, S., Olofson, K., Andersson, P., Pettersson, J., and Hallquist, M. (2006) *Evolution of the urban aerosol during winter temperature inversion episodes*. *Atmos. Environ.*, **40**: 5355–5366, DOI:10.1016/j.atmosenv.2006.04.051
- Kanda, M. (2007) *Progress in urban meteorology: A review*. *J. Meteor. Soc. Japan*, 85B, 363–383, DOI:10.2151/jmsj.85B.363
- Lee, S. H. & Kim, H. D. (2008) *Effects of regional warming due to urbanization on daytime local circulations in a complex basin of the Daegu metropolitan area, Korea*. *J. Appl. Meteor. Climatol.*, **47**:1427–1441, DOI:10.1175/JAMC1504.1
- Lidsky, T.I. & Schneider, J. S. (2006) *Adverse Effects of Childhood Lead Poisoning: The Clinical Neuropsychological Perspective*. 100:2 *Env. Research* 284
- Panday, A. & Prinn, R. (2009) *Diurnal cycle of air pollution in the Kathmandu Valley, Nepal: Observations*. *J. Geophys. Res.*, 114, DOI: 10.1029/2008JD009777
- Perez-Foguet, A. (2014) *Characterization of local wind patterns in complex mountain valleys*. *International Journal of Climatology*. **34**: 1741-1759. DOI: 10.1002/joc.3798
- Pebe, G., Villa, H., Escate, L., Cervantes, G. (2008) *Blood lead levels in newborns from La Oroya, 2004–2005*. *Rev Peru Med Exp Salud Publica*; 25:355–359
- R Core Team (2016) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL: <https://www.R-project.org/>
- Ratto, G. E., Videla, F., Maronna, R., Flores, A., Davila, F. P. (2010) *Air pollutant transport analysis based on hourly winds in the city of La Plata and surroundings, Argentina*. *Water, Air, and Soil Pollution* **208**: 243-257. DOI: 10.1007/s11270-009-0163-0
- Raaschou-Nielsen, O (2013) *Air pollution and lung cancer incidence in 17 European cohorts: prospective analyses from the European Study of Cohorts for Air Pollution Effects (ESCAPE)*. *The Lancet Oncology* 14 (9): 813–22. DOI:10.1016/S1470-2045(13)70279-1 . PMID 23849838
- Rizwan, A., Dennis, L., and Liu, C. (2008) *A review on the generation, determination and mitigation of urban heat island*, *J. Environ. Sci. (China)*, **20**, 120–128, doi:10.1016/S1001-0742(08)60019-4
- Romero, H., Rivera, M., Zalazar, P., and Azocar, P. (1999) *Rapid urban growth, land-use changes and air pollution in Santiago, Chile*. *Atmos. Environ.* **33**, 4039–4047, DOI:10.1016/S1352-2310(99)00145-4
- Savijärvi, H., & Liya, J. (2001) *Local winds in a valley city*, *Bound.-Layer Meteor.*, **100**, 301–319, DOI:10.1023/A:1019215031007
- Scutari, M. & Denis, J-B. (2014) *Bayesian Networks with Examples in R*. Chapman and Hall, Boca Raton. ISBN 978-1482225587
- Valencia, A. (2012) *Human rights trade-offs in a context of systemic unfreedom: The case of the smelter town of La Oroya, Peru*, The Centre for Development Studies University of Bath, Claverton Down Bath, BA2 7AY, UK, ISSN 2040-3151
- Yao, W. & Zhong, S. (2009) *Nocturnal temperature inversions in a small, enclosed basin and their relationship to ambient atmospheric conditions*. *Meteor. Atmos. Phys.*, **103**, 195–210, DOI:10.1007/s00703-008-0341-4

United States Securities and Exchange Commission (2001) *FORM 10-K: Doe Run Corporation*. ANNUAL REPORT PURSUANT TO SECTION 13 OR 15(d) OF THE SECURITIES AND EXCHANGE ACT OF 1934.

United States Securities and Exchange Commission. Available from:

<http://www.sec.gov/Archives/edgar/data/1061112/000091205702021037/a2079883z10-k.txt>

[Accessed: 28/04/2016]

United States Securities and Exchange Commission (2006) *FORM 10-K: Doe Run Corporation*. ANNUAL REPORT PURSUANT TO SECTION 13 OR 15(d) OF THE SECURITIES AND EXCHANGE ACT OF 1934.

United States Securities and Exchange Commission. Available from:

[http://www.sec.gov/Archives/edgar/data/1061112/000110465906018264/a06-5938\\_110k.htm](http://www.sec.gov/Archives/edgar/data/1061112/000110465906018264/a06-5938_110k.htm)

[Accessed: 28/04/2016]